

-1-

-2-

Human Proteins

Field of the Invention

The present invention relates to genes encoding novel human proteins which exhibit a variety useful biological activities. More specifically, isolated nucleic acid molecules are provided which encode polypeptides comprising various forms of human proteins. Human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are methods for detecting nucleic acids or polypeptides related to those of the invention, for example, to aid in identification of a biological sample or diagnosis of disorders related to expression of protein genes of this invention. The invention further relates to methods for identifying agonists and antagonists of the proteins of the invention, as well as to methods for treatment of disorders related to protein gene expression using polypeptides, antagonists and agonists of the invention.

Background of the Invention

Identification and sequencing of human genes is a major goal of modern scientific research. For example, by identifying genes and determining their sequences, scientists have been able to make large quantities of valuable human gene products. These include human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, erythropoietin and numerous other proteins. Additionally, knowledge of gene sequences can provide keys to diagnosis, treatment or cure of genetic diseases such as muscular dystrophy and cystic fibrosis.

Despite the great progress that has been made in recent years, only a small number of genes which encode the presumably thousands of human proteins have been identified and sequenced. Therefore, there is a need for identification and characterization of novel human proteins and corresponding genes which can play a role in detecting, preventing, ameliorating or correcting disorders related to abnormal expression of and responses to such proteins.

Summary of the Invention

The present invention provides isolated nucleic acid molecules comprising polynucleotide sequences which have been identified as sequences encoding human proteins of the invention. Each protein of the invention is identified in Table 1, below (see Example 2) by a reference number designated as a "Protein ID (Identifier)" (e.g., "PF353-01"). Each protein of the invention is related to a human complementary DNA (cDNA) clone prepared from a messenger RNA (mRNA) encoding the related protein. The cDNA clone related to each protein of the invention is identified by a "cDNA Clone ID (Identifier)" in Table 1 (e.g., "HABCE99"). DNA of each cDNA clone in Table 1 is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown for each cDNA Clone ID in Table 1, as further described below.

The invention provides a nucleotide sequence determined for an mRNA molecule encoding each protein identified in Table 1, which is designated in Table 1 as the "Total NT (Nucleotide Sequence)." This determined nucleotide sequence has been assigned a SEQ ID NO = "X" in the Sequence Listing hereinbelow, where the value of X for the determined nucleotide sequence of each protein is an integer specified in Table 1. The determined nucleotide sequence provided for each protein of the invention was determined by applying conventional automated nucleotide sequencing methods to DNA of the corresponding deposited cDNA clone cited in Table 1.

The determined nucleotide sequence for the mRNA encoding each protein of the invention has been translated to provide a determined amino acid sequence for each protein which is identified in Table 1 by a SEQ ID NO = "Y" where the value of Y for each protein is an integer defined in Table 1. The determined amino acid sequence for each protein represents the amino acid sequence encoded by the determined nucleotide sequence, beginning at or near the translation initiation ("start") codon of the protein and continuing until the first translation termination ("stop") codon. Due to possible errors inherent in determining nucleotide sequences from any DNA molecule, particularly using the conventional automated sequencing technology used

-3-

-4-

to sequence the cDNA clones described herein, occasional nucleotide sequence errors are expected in the determined nucleotide sequences of the invention. These errors may include insertions or deletions of one or a few nucleotides in the determined nucleotide sequence as compared to the actual nucleotide sequence of the deposited cDNA. As one of ordinary skill would appreciate, incorrect insertions or deletions of one or two nucleotides into a determined nucleotide sequence leads to a shift in the translation reading frame compared to the reading frame actually encoded by a cDNA clone. Further, such a shift in frame within an actual open reading frame frequently leads to the appearance of a translation termination (stop) codon within the sequence encoding the polypeptide. Accordingly, due to occasional errors in the nucleotide sequences determined from the deposited cDNAs and any related DNA clones used to prepare the determined sequence for the mRNA encoding each secreted protein of the invention, the translations shown as determined amino acid sequences in SEQ ID NO:Y may represent only a portion of the complete amino acid sequence of the human secreted protein actually encoded by the mRNA represented by the corresponding cDNA clone in the ATCC deposit identified in Table 1. In any event, the determined amino acid sequence for each protein in Table 1, which is shown in SEQ ID NO:Y for each protein, comprises at least a portion of the amino acid sequence determined for that protein.

20 More particularly, the determined amino acid sequence is the amino acid sequence translated from the determined nucleotide sequence in the open reading frame of the first amino acid of the ORF to the last amino acid of that frame. In other words, the determined amino acid sequence is translated from the determined nucleotide sequence beginning at the codon having as its 5' end the nucleotide in the position of SEQ ID NO:X identified in Table 1 as the 5' nucleotide of the first amino acid (abbreviated in Table 1 as "5' NT of First AA"). Translation of the determined nucleotide sequence is continued in the reading frame of that first amino acid codon to the first stop codon in that same open reading frame, i.e., to the position in SEQ ID

5 one or two nucleotides into a determined nucleotide sequence leads to a shift in the translation reading frame compared to the reading frame actually encoded by a cDNA clone. Further, such a shift in frame within an actual open reading frame frequently leads to the appearance of a translation termination (stop) codon within the sequence encoding the polypeptide. Accordingly, due to occasional errors in the nucleotide sequences determined from the deposited cDNAs and any related DNA clones used to prepare the determined sequence for the mRNA encoding each secreted protein of the invention, the translations shown as determined amino acid sequences in SEQ ID NO:Y may represent only a portion of the complete amino acid sequence of the human secreted protein actually encoded by the mRNA represented by the corresponding cDNA clone in the ATCC deposit identified in Table 1. In any event, the determined amino acid sequence for each protein in Table 1, which is shown in SEQ ID NO:Y for each protein, comprises at least a portion of the amino acid sequence determined for that protein.

10 For any determined amino acid sequence in which the first amino acid is the methionine encoded by the translation initiation codon for the protein, Table 1 also identifies the position in SEQ ID NO:X of the 5' nucleotide of the start codon ("5' NT of Start Codon") as the same position in SEQ ID NO:X as that of the 5' nucleotide of the first amino acid ("First AA").

15 Table 1 also identifies the positions in SEQ ID NO:Y of the last amino acid of the signal peptide ("Last AA of Sig Pep") and the first amino acid of the secreted portion ("First AA of Secreted Portion") of the protein, for those polypeptides having a secretory leader sequence. The "secreted portion" of a secreted protein in the present context indicates that portion of the complete polypeptide translated from an mRNA which remains after cleavage of the signal peptide by a signal peptidase. In this context the term "mature" may also be used interchangeably with "secreted portion" although it is recognized that in other contexts "mature" may designate a portion of a "proprotein" which is produced by further cleavage of the polypeptide after cleavage of the signal peptide.

20 Accordingly, in one aspect the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence which is identical to the nucleotide sequence of SEQ ID NO:X, where X is any integer as defined in Table 1. The invention also provides an isolated nucleic acid molecule comprising a nucleotide sequence which is identical to a portion of the nucleotide sequence of SEQ ID NO:X, for instance, a sequence of at least 50, 100 or 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X. Such a portion of the nucleotide sequence of SEQ ID NO:X may be described most generally as a sequence of at least C contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X where: (1) the sequence of at least C contiguous nucleotides begins with the nucleotide at position N of SEQ ID NO:X and ends with the nucleotide at position M of SEQ ID NO:X; (2) C is any integer in the range beginning with a convenient primer size, for instance, about 20, to

NO:X which encodes the amino acid at the position in SEQ ID NO:Y identified as the "last amino acid of the open reading frame" (abbreviated as "Last AA of ORF").

For any determined amino acid sequence in which the first amino acid is the methionine encoded by the translation initiation codon for the protein, Table 1 also identifies the position in SEQ ID NO:X of the 5' nucleotide of the start codon ("5' NT of Start Codon") as the same position in SEQ ID NO:X as that of the 5' nucleotide of the first amino acid ("First AA").

Table 1 also identifies the positions in SEQ ID NO:Y of the last amino acid of the signal peptide ("Last AA of Sig Pep") and the first amino acid of the secreted portion ("First AA of Secreted Portion") of the protein, for those polypeptides having a secretory leader sequence. The "secreted portion" of a secreted protein in the present context indicates that portion of the complete polypeptide translated from an mRNA which remains after cleavage of the signal peptide by a signal peptidase. In this context the term "mature" may also be used interchangeably with "secreted portion" although it is recognized that in other contexts "mature" may designate a portion of a "proprotein" which is produced by further cleavage of the polypeptide after cleavage of the signal peptide.

Accordingly, in one aspect the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence which is identical to the nucleotide sequence of SEQ ID NO:X, where X is any integer as defined in Table 1. The invention also provides an isolated nucleic acid molecule comprising a nucleotide sequence which is identical to a portion of the nucleotide sequence of SEQ ID NO:X, for instance, a sequence of at least 50, 100 or 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X. Such a portion of the nucleotide sequence of SEQ ID NO:X may be described most generally as a sequence of at least C contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X where: (1) the sequence of at least C contiguous nucleotides begins with the nucleotide at position N of SEQ ID NO:X and ends with the nucleotide at position M of SEQ ID NO:X; (2) C is any integer in the range beginning with a convenient primer size, for instance, about 20, to

-5-

the total nucleotide sequence length ("Total NT Seq.") as set forth for SEQ ID NO:X in Table 1; (3) N is any integer in the range of 1 to the first position of the last C nucleotides in SEQ ID NO:X, or more particularly, N is equal to the value of Total NT Seq. minus the quantity C plus 1 (i.e., Total NT Seq.-(C+1)); and (4) M is any integer in the range of C to Total NT Seq.

Preferably, the sequence of contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X is included in SEQ ID NO:X in the range of positions beginning with the nucleotide at about the 5' nucleotide of the clone sequence ("5' NT of Clone Seq." in Table 1) and ending with the nucleotide at about the 3' nucleotide of the clone sequence ("3' NT of Clone Seq." in Table 1). More preferably, the sequence of contiguous nucleotides is in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon ("5' NT of Start Codon" in Table 1) and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as set forth for SEQ ID NO:X in Table 1. For instance, one preferred embodiment of this aspect of the invention is an isolated nucleic acid molecule which comprises a sequence at least 95%, 96%, 97%, 98%, or 99% identical to a sequence of about 500 contiguous nucleotides included in the nucleotide sequence of SEQ ID NO:X beginning at about the 5' NT of Start Codon position as set forth for SEQ ID NO:X in Table 1. Another preferred embodiment of this aspect of the invention is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Further embodiments of the invention include isolated nucleic acid molecules which comprise a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98%, 99% or 99.9% identical, to any of the determined nucleotide sequences above. For instance, one such embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a

sequence of at least 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1. Another embodiment of this aspect of the invention is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Isolated nucleic acid molecules which hybridize under stringent hybridization conditions to a nucleic acid molecule described above also are provided. Such a nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

The invention further provides a composition of matter comprising a nucleic acid molecule which comprises a human cDNA clone identified by a cDNA Clone ID (Identifier) in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for that cDNA clone. As described further in Example 1, this deposited material comprises a mixture of plasmid DNA molecules containing cloned cDNAs of the invention. Further, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence which is, for instance, at least 95% identical to a sequence of at least 50, 150 or 500 contiguous nucleotides in the nucleotide sequence encoded by a human cDNA clone contained in the deposit given the ATCC Deposit Number shown in Table 1. One preferred embodiment of this aspect is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by a human cDNA clone identified in Table 1 and as contained in the deposit with the ATCC Deposit Number shown in Table 1. Also provided are isolated nucleic acid molecules which hybridize under stringent hybridization conditions to a nucleic acid molecule comprising a nucleotide sequence encoded by a human cDNA clone identified in Table 1 and contained in the cited deposit.

-6-

These nucleic acid molecules of the invention may be used for a variety of identification and diagnostic purposes. For instance, the invention provides a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a nucleotide sequence of the invention. The sequence of the nucleic acid molecule used in this method is selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. This method of the invention comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in the biological sample with a sequence selected from the group above, and determining whether the sequence of the nucleic acid molecule in the sample is at least 95% identical to the selected sequence. The step of comparing sequences may comprise determining the extent of nucleic acid hybridization between nucleic acid molecules in the sample and a nucleic acid molecule comprising the sequence selected from the above group. Alternatively, this step may be performed by comparing the nucleotide sequence determined from a nucleic acid molecule in the sample, for instance by automated DNA sequence methods, with the sequence selected from the above group.

In another aspect, the invention provides methods for identifying the species, tissue or cell type of a biological sample based on detecting nucleic acid molecules in the sample which comprise a nucleotide sequence of a nucleic acid molecule of the invention (for instance, a nucleic acid molecule comprising a nucleotide sequence that is at least 95% identical to at least a portion of a nucleotide sequence of SEQ ID NO:X or a nucleotide sequence encoded by a human cDNA clone identified in Table 1 as contained in the deposit with the ATCC Deposit Number shown therein. This method may be conducted by detecting a nucleotide sequence of an individual cDNA of the invention or using panel of nucleotide sequences of the invention. Thus, this method may comprise a step of detecting nucleic acid molecules comprising a

nucleotide sequence in a panel of at least two nucleotide sequences, where at least one sequence in the panel is at least 95% identical to at least a portion of a nucleotide sequence of SEQ ID NO:X or a nucleotide sequence encoded by a human cDNA clone contained in the ATCC deposit. In this method for identifying the species, tissue or cell type of a biological sample, the detection of nucleic acid molecules comprising nucleotide sequences of the invention may be conducted by various techniques known in the art including, for instance, hybridization of either DNA or RNA probes to either DNA or RNA molecules obtained from the biological sample, as well as computational comparisons of nucleotide sequences determined from nucleic acids in a biological sample with nucleotide sequences of the invention.

Similarly, nucleic acid molecules of the invention may be used in a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a protein identified in Table 1. This method may comprise a step of detecting in a biological sample obtained from the subject nucleic acid molecules comprising a nucleotide sequence that is at least 95% identical to at least a portion of a nucleotide sequence of SEQ ID NO:X or a nucleotide sequence encoded by a human cDNA clone identified in Table 1 as contained in the deposit with the given ATCC Deposit Number. Again, this diagnostic method may involve analysis of individual nucleotide sequences or panels of several nucleotide sequences, and the analysis of either DNA or RNA species using either DNA or RNA probes. For use in identification or diagnostic methods such as those described above, therefore, the invention also provides a composition of matter comprising isolated nucleic acid molecules in which the nucleotide sequences of the nucleic acid molecules comprise a panel of sequences, at least one of which is at least 95% identical to a sequence, either a nucleotide sequence of SEQ ID NO:X or a nucleotide sequence encoded by a human cDNA clone contained in the ATCC deposit in Table 1. In this composition, the nucleic acid molecules may comprise DNA molecules or RNA molecules or both, as well as polynucleotide equivalents of DNA and RNA which are not naturally occurring but are known in the art as such.

Another aspect of the invention relates to polypeptides comprising amino acid sequences encoded by nucleotide sequences of the invention. For identification and diagnostic purposes, these polypeptides need not include the amino acid sequence of a complete secreted protein or even of the secreted form of such a protein, since, for instance, antibodies may bind specifically to a linear epitope of a polypeptide which comprises as few as 6 to 8 amino acids. Accordingly, the invention also provides an isolated polypeptide comprising an amino acid sequence at least 90%, preferably 95%, 96%, 97%, 98%, or 99% identical to a sequence of at least about 10, 30 or 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1. Preferably, the sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y beginning with the residue at about the position of the First Amino Acid of the Secreted Portion where one exists or the first amino acid of the open reading frame if the protein is not indicated as having a signal peptide and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1. A preferred embodiment of this aspect relates to an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

As noted above, however, the determined amino acid sequence of SEQ ID NO:Y may not include the complete amino acid sequence of the protein encoded by each cDNA in the ATCC deposit identified in Table 1. Accordingly, the invention further provides an isolated polypeptide comprising an amino acid sequence at least 90% identical, preferably at least 95%, 96%, 97%, 98% or 99% identical to a sequence of at least about 10, 300 or 100 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for that cDNA clone in Table 1. A particularly preferred embodiment of this aspect is a polypeptide in which the sequence of contiguous amino acids is included in the amino acid sequence of a secreted ("mature") portion of the protein encoded by a human cDNA clone contained in the deposit, particularly a

polypeptide comprising the entire amino acid sequence of the secreted portion of the secreted protein encoded by a human cDNA clone of the invention.

For purposes such as tissue identification and diagnosis of pathological conditions, the invention also provides an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence of the invention, (for instance, a sequence that is identical to a sequence of at least 6, preferably at least 7, 8, 9 or 10, contiguous amino acids in an amino acid sequence of SEQ ID NO:Y or in a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit cited therein. Further in the same vein, the invention provides a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is identical to a sequence of at least 6, preferably at least 7, 8, 9 or 10 contiguous amino acids in a sequence selected from the group consisting of an amino acid sequence of SEQ ID NO:Y and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for that cDNA clone in Table 1. This method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from the above group and determining whether the sequence of that polypeptide molecule in the sample is identical to the selected sequence of at least 6-10 contiguous amino acids. This step of comparing an amino acid sequence of at least one polypeptide molecule in the sample with a sequence selected from the above group may comprise determining the extent of specific binding of polypeptides in the sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence of the invention. Alternatively, this comparison step may be performed by comparing the amino acid sequence determined from a polypeptide molecule in the sample with the sequence selected from the above group, for instance, using computational methods.

The invention further provides methods for identifying the species, tissue or cell type of a biological sample comprising a step of detecting polypeptide molecules

-11-

-12-

in the sample which include an amino acid sequence that is identical to a sequence of at least 6-10 contiguous amino acids an amino acid sequence of SEQ ID NO:Y or of a cDNA identified in Table 1 and contained in the cited deposit. This method may involve analyses of polypeptides for the presence of individual amino acid sequences of the invention or of panels of such sequences. Similarly provided are methods for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a protein identified in Table 1. In preferred embodiments of these methods of the invention for identification or diagnosis, an antibody which binds specifically to a polypeptide comprising an amino acid sequence of the invention is used to analyze amino acid sequences of polypeptides in a biological sample.

In yet another aspect, the invention provides recombinant means for making a polypeptide comprising all or a portion of an amino acid sequence of the invention. For this purpose, an isolated nucleic acid molecule comprising a nucleotide sequence which is, for instance, at least 95% identical to a nucleotide sequence encoding a polypeptide which comprises an amino acid sequence of the invention (for instance, one that is at least 90% identical to SEQ ID NO:Y.

It will be readily appreciated by one of ordinary skill that, due to the degeneracy of the genetic code, any nucleotide sequence encoding the amino acid sequence of a given protein needs to share only a low level of identity with the nucleotide sequence of a human cDNA clone which encodes the identical amino acid sequence of that protein. It will be further appreciated that the nucleotide of the deposited cDNAs presumably all comprise codons optimized for expression by human cells from which the cDNAs originated. Therefore, for improved expression in recombinant prokaryotic host cells, for instance, it may be desirable to alter the codon usage in a nucleic acid molecule encoding an amino acid sequence of the invention, selecting codons in accordance with the redundancy of the genetic code, which provide optimal codon usage in the selected host. Preferred nucleic acid molecules of this aspect of the invention are those which encode a polypeptide which comprises an

complete amino acid sequence of SEQ ID NO:Y or a complete amino acid sequence of a protein encoded by a human cDNA clone identified in Table 1 and contained in the deposit cited therein.

Using such nucleic acid molecules encoding polypeptides of the invention, the invention further provides recombinant means for making the polypeptides. Thus, included is a method of making a recombinant vector comprising inserting an isolated nucleic acid molecule of the invention into a vector, as well as a recombinant vector produced by this method. Also included is a method of making a recombinant host cell comprising introducing a vector of the invention into a host cell, and a recombinant host so made. Such cells are useful, for instance, in a method of making an isolated polypeptide of the invention which comprises culturing a recombinant host cell under conditions such that the polypeptide is expressed and recovering the polypeptide.

In a preferred embodiment of this method, the recombinant host cell is a eukaryotic cell and the polypeptide encoded by the nucleic acid of the invention encodes the complete amino acid sequence of a protein encoded by a cDNA identified in Table 1, so that the polypeptide produced by this method is a secreted ("mature") portion of a human secreted protein of the invention (i.e., one comprising an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position identified in Table 1 as the First AA of Secreted Portion of SEQ ID NO:Y or an amino acid sequence of a secreted portion of a secreted protein encoded by a human cDNA clone identified in Table 1 and contained in the deposit with the ATCC Deposit Number shown in Table 1. The invention further provides an isolated polypeptide which is a secreted portion of a human secreted protein produced by the above method. Where the polypeptide shown in Table 1 does not have a leader sequence one may be provided by the vector. Such vectors are known in the art and are discussed below.

In yet another aspect, the invention provides a method of treatment of an individual in need of an increased level of a secreted protein activity. As described herein, diagnostic methods of the invention enable the identification of such individuals, that is, individuals with a pathological condition involving a particular

-13-

organ, tissue or cell type, exhibiting lower levels of expression product (e.g., mRNA or antigen) of a given secreted protein in that organ, tissue or cell type, or those with mutant expression products, compared with normal individuals not suffering from the pathology. The method of the invention for treatment of an individual with such a pathological condition comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide of a secreted protein of the invention effective to increase the level of activity of that secreted protein in the individual.

Agonists and antagonists of the polypeptides of the invention and methods for using these also are provided.

Brief Description of the Drawings

Figure 1 shows the nucleotide sequence and deduced amino acid sequence of CCV (HEMF185), SEQ ID NOS:1 and 2, respectively.

Figure 2 shows the nucleotide sequence and deduced amino acid sequence of CAT-1 (HTXET53), SEQ ID NOS:3 and 4, respectively.

Figure 3 shows the nucleotide sequence and deduced amino acid sequence of CAT-2 (HT3SG28), SEQ ID NOS:5 and 6, respectively.

Figure 4 shows the nucleotide sequence and deduced amino acid sequence of MIA-2 (HBXAK03), SEQ ID NOS:7 and 8, respectively.

Figure 5 shows the nucleotide sequence and deduced amino acid sequence of MIA-3 (HLFBD44), SEQ ID NOS:9 and 10, respectively.

Figure 6 shows the nucleotide sequence and deduced amino acid sequence of AIF-2 (HEBGM49), SEQ ID NOS:11 and 12, respectively.

Figure 7 shows the nucleotide sequence and deduced amino acid sequence of AIF-3 (HNGBH45), SEQ ID NOS:13 and 14, respectively.

Figure 8 shows the nucleotide sequence and deduced amino acid sequence of Annexin (HSAAL25), SEQ ID NOS:15 and 16, respectively.

Figure 9 shows the nucleotide sequence and deduced amino acid sequence of ES130-like I (HUSAX55), SEQ ID NOS:17 and 18, respectively.

Figure 10 shows the nucleotide sequence and deduced amino acid sequence of BEF (HSXCK41), SEQ ID NOS:19 and 20, respectively.

Figure 11 shows the nucleotide sequence and deduced amino acid sequence of ADF (HFKFY79), SEQ ID NOS:21 and 22, respectively.

Figure 12 shows the nucleotide sequence and deduced amino acid sequence of Bcl-like (HAICH28), SEQ ID NOS:23 and 24, respectively.

Detailed Description

Nucleic Acid Molecules

Human Proteins

The present invention provides isolated nucleic acid molecules comprising polynucleotide sequences which have been identified as sequences encoding human proteins. The invention further provides a nucleotide sequence determined from an mRNA molecule encoding each human protein identified in Table 1, which comprises all or a substantial portion of the complete nucleotide sequence of the mRNA encoding each protein of the invention and has been assigned a SEQ ID NO = "X" in the Sequence Listing and Figures hereinbelow.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring nucleic acid molecule or polynucleotide present in a living organism is not isolated, but the same nucleic acid molecule or polynucleotide, separated from some or all of the coexisting materials in the natural environment, is isolated. Such nucleic acid molecule could be part of a vector and/or such polynucleotide could be part of a composition, and still be isolated in that such vector or composition is not part of the natural environment of the nucleic acid molecule or polynucleotide.

By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides,

-14-

Figure 10 shows the nucleotide sequence and deduced amino acid sequence of BEF (HSXCK41), SEQ ID NOS:19 and 20, respectively.

Figure 11 shows the nucleotide sequence and deduced amino acid sequence of ADF (HFKFY79), SEQ ID NOS:21 and 22, respectively.

Figure 12 shows the nucleotide sequence and deduced amino acid sequence of Bcl-like (HAICH28), SEQ ID NOS:23 and 24, respectively.

Detailed Description

Nucleic Acid Molecules

Human Proteins

The present invention provides isolated nucleic acid molecules comprising polynucleotide sequences which have been identified as sequences encoding human proteins. The invention further provides a nucleotide sequence determined from an mRNA molecule encoding each human protein identified in Table 1, which comprises all or a substantial portion of the complete nucleotide sequence of the mRNA encoding each protein of the invention and has been assigned a SEQ ID NO = "X" in the Sequence Listing and Figures hereinbelow.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring nucleic acid molecule or polynucleotide present in a living organism is not isolated, but the same nucleic acid molecule or polynucleotide, separated from some or all of the coexisting materials in the natural environment, is isolated. Such nucleic acid molecule could be part of a vector and/or such polynucleotide could be part of a composition, and still be isolated in that such vector or composition is not part of the natural environment of the nucleic acid molecule or polynucleotide.

By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides,

-15-

-16-

and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U).

Using the information provided herein, such as a nucleotide sequence shown in the sequence listing, a nucleic acid molecule of the present invention encoding a polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. The present invention provides not only the determined nucleotide sequences of the mRNA encoding each human secreted protein of the invention, as set forth in SEQ ID NO:X for each

protein, but also a sample of plasmid DNA containing a cDNA of the invention deposited with the American Type Culture Collection (Rockville, MD), as set forth in Table 1. These deposits enable recovery of each cDNA clone and recombinant production of each secreted protein of the invention actually encoded by a cDNA clone identified in Table 1, as further described hereinbelow.

10 Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

20

In addition to nucleic acid molecules comprising a determined nucleotide sequence in SEQ ID NO:X or the nucleotide sequence of a deposited human cDNA clone, isolated nucleic acid molecules of the invention include DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the proteins shown in the sequence listing or those encoded by the clones contained in the deposited plasmids. Of course, the genetic code and species-specific codon preferences are well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g.,

change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*). Preferably, this nucleic acid molecule will encode a secreted portion (nature polypeptide) encoded by the deposited cDNA.

The invention further provides a nucleic acid molecule having a sequence 5 complementary to one of the above sequences. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, *in situ* hybridization with chromosomes, and for detecting expression of the corresponding gene(s) in human tissue, for instance, by Northern blot analysis.

The present invention is further directed to nucleic acid molecules encoding 10 portions of the nucleotide sequences described herein as well as to fragments of the isolated nucleic acid molecules described herein. By a "fragment" of an isolated nucleic acid molecule having the nucleotide sequence of the deposited cDNA or the nucleotide sequence shown in the sequence listing is intended fragments at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50-500 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the deposited cDNA or as shown in the sequence listing. By a fragment "at least 20 nt in length," for example, is intended fragments which include 20 or more contiguous bases from the nucleotide

20 sequence of the deposited cDNA or the determined nucleotide sequence shown in SEQ ID NO:X. Preferred nucleic acid fragments of the present invention include nucleic acid molecules encoding epitope-bearing portions of the polypeptides of the present invention, as described further below.

25 In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of a nucleic acid molecule of the invention described above, for instance, a cDNA contained in the plasmid sample deposited with the ATCC. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a

-17-

solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μ g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 (e.g., 50) nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below. For certain applications, such as the FISH technique for gene mapping on chromosomes, probes of 500 nucleotides up to 2000 nucleotides may be preferred.

By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., the deposited cDNA or the nucleotide sequence as shown in SEQ ID NC:X). Of course, a polynucleotide which hybridizes only to a poly A sequence (such as any 3' terminal poly(A) tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

Also encoded by nucleic acids of the invention are the amino acid sequences of the invention together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing, including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; and additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities.

Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell* 37: 767 (1984). As discussed below, other such fusion proteins include those fused to Fc at the N- or C-terminus.

Sequences Encoding Signal Peptide and Secreted Portions

According to the signal hypothesis, proteins secreted by eukaryotic cells have a signal peptide (or secretory leader sequence) which is cleaved from the complete polypeptide to produce a secreted portion or "mature" form of the protein. Methods for predicting whether a protein has a signal peptide (or "secretory leader") as well as the cleavage point for that leader sequence are well known in the art. See, for instance, von Heijne, *supra*. The determined amino acid sequence of several proteins of the invention, determined by translation of the determined nucleotide sequence identified in Table 1, have been found to comprise an amino acid sequence of a secretory signal peptide. The sequence and cleavage site of that signal peptide are described in Table 1 and in the Examples and the signal sequence is underlined in the Figures, to the extent that these have been determined for each secreted protein of the invention.

More in particular, the present invention provides nucleic acid molecules encoding a secreted portion (mature form) of each secreted protein identified in Table 1.

1. Most mammalian cells and even insect cells cleave signal peptides from secreted proteins with approximately the same specificity. However, in some cases, cleavage

-18-

Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell* 37: 767 (1984). As discussed below, other such fusion proteins include those fused to Fc at the N- or C-terminus.

According to the signal hypothesis, proteins secreted by eukaryotic cells have a signal peptide (or secretory leader sequence) which is cleaved from the complete polypeptide to produce a secreted portion or "mature" form of the protein. Methods for predicting whether a protein has a signal peptide (or "secretory leader") as well as the cleavage point for that leader sequence are well known in the art. See, for instance, von Heijne, *supra*. The determined amino acid sequence of several proteins of the invention, determined by translation of the determined nucleotide sequence identified in Table 1, have been found to comprise an amino acid sequence of a secretory signal peptide. The sequence and cleavage site of that signal peptide are described in Table 1 and in the Examples and the signal sequence is underlined in the Figures, to the extent that these have been determined for each secreted protein of the invention.

More in particular, the present invention provides nucleic acid molecules encoding a secreted portion (mature form) of each secreted protein identified in Table 1.

1. Most mammalian cells and even insect cells cleave signal peptides from secreted proteins with approximately the same specificity. However, in some cases, cleavage

Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell* 37: 767 (1984). As discussed below, other such fusion proteins include those fused to Fc at the N- or C-terminus.

According to the signal hypothesis, proteins secreted by eukaryotic cells have a signal peptide (or secretory leader sequence) which is cleaved from the complete polypeptide to produce a secreted portion or "mature" form of the protein. Methods for predicting whether a protein has a signal peptide (or "secretory leader") as well as the cleavage point for that leader sequence are well known in the art. See, for instance, von Heijne, *supra*. The determined amino acid sequence of several proteins of the invention, determined by translation of the determined nucleotide sequence identified in Table 1, have been found to comprise an amino acid sequence of a secretory signal peptide. The sequence and cleavage site of that signal peptide are described in Table 1 and in the Examples and the signal sequence is underlined in the Figures, to the extent that these have been determined for each secreted protein of the invention.

More in particular, the present invention provides nucleic acid molecules encoding a secreted portion (mature form) of each secreted protein identified in Table 1.

1. Most mammalian cells and even insect cells cleave signal peptides from secreted proteins with approximately the same specificity. However, in some cases, cleavage

-19-

-20-

of the signal peptide (as referred to herein as a "leader sequence" or "leader") from a secreted protein is not entirely uniform, which results in more than one secreted (also herein "mature") form or species of the protein. Further, it has long been known that the cleavage specificity of a secreted protein is ultimately determined by the primary

structure of the complete protein, that is, it is inherent in the amino acid sequence of the initial polypeptide translated from its mRNA. Therefore, the present invention provides not only a determined nucleotide sequence and translated amino acid sequence identifying a signal peptide and secreted portion of each secreted protein of the invention, but also a deposited sample of a cDNA clone encoding a secreted (mature) form of each secreted protein of the invention.

More particularly, the invention further provides an isolated polypeptide comprising an amino acid sequence at least 90% identical, preferably 95%, 96%, 97%, 98% or 99% identical, to a sequence of at least about 25, 50 or 100 contiguous amino acids in the complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for that cDNA clone in Table 1. A particularly preferred embodiment of this aspect of the invention is a polypeptide in which the sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. By the "secreted portion [or mature form]" of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1" is meant the secreted

portion(s) or mature form(s) of the protein produced by expression in any eukaryotic cell (for instance, cells of an established insect or mammalian cell line), preferably a human cell (for instance, cells of the well known HeLa cell line), of the complete open reading frame encoded by the human cDNA clone identified in Table 1 and contained in the deposit cited in Table 1.

Variant and Mutant Polynucleotides

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the secreted proteins. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the secreted protein or portions thereof. Also especially preferred in this regard are conservative substitutions.

Most highly preferred are nucleic acid molecules encoding a secreted portion (mature form) of a protein described in Table 1 and having the amino acid sequence shown in the sequence listing as SEQ ID NO:X, or the amino acid sequence of the secreted portion (mature form) of the protein encoded by a deposited cDNA clone. Further embodiments include an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 85% identical, more preferably at least 90% identical, and most preferably at least 95%, 96%, 97%, 98% or 99% identical to a polynucleotide of the invention described in Table 1, or a polynucleotide which hybridizes under stringent hybridization conditions to such a polynucleotide. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide

-21-

which encodes the amino acid sequence of an epitope-bearing portion of a secreted polypeptide having an amino acid sequence of SEQ ID NO:Y or an amino acid sequence of a secreted protein encoded by a cDNA clone in the deposit identified in Table 1.

5 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a secreted polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the secreted polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in SEQ ID NO:1, or to the nucleotide sequence of a deposited cDNA can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such

10 15 20 25

that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

Table 1. Uses for Nucleic Acid Molecules of the Invention

5 Each of the nucleic acid molecules identified herein can be used in numerous ways as polynucleotide reagents. The polynucleotides can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type. In addition, these polynucleotides can be used as diagnostic probes suitable for use in genetic linkage analysis (polymorphisms). Further, the polynucleotides can be used as probes for locating gene regions associated with genetic disease, as explained in more detail below.

The polynucleotides of the present invention are also valuable for chromosome identification. Each polynucleotide is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of cDNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in the sequence listing. Computer analysis of the sequences is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the secreted protein will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular nucleic acid sequence to a particular chromosome. Three or more clones can

-23-

-24-

be assigned per day using a single thermal cycler. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map a gene to its

chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-

cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase

chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique

chromosomal location with sufficient signal intensity for simple detection. For

example, 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this

technique, see Verna et al., *Human Chromosomes: a Manual of Basic Techniques*. Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a polynucleotide sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library.) The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR, based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

In addition to the foregoing, the polynucleotides of the invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, *Nucl. Acids Res.*, 6:3073 (1979); Cooney et al, *Science*, 241:456 (1988); and Dervan et al, *Science*, 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.*, 56:560 (1991) *Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

-25-

Nucleic acid molecules of the present invention are also a useful in gene therapy which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect. The high specificity of the cDNA probes according to this invention offer means of targeting such gene locations in a highly accurate manner.

5 The sequences of the present invention, as broadly defined, are also useful for identification of individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP.

10 However, RFLP is a pattern based technique, which does not require the DNA sequence of the individual to be sequenced. The polynucleotides and sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take a sequence of the invention and prepare two PCR primers. These are used to amplify an individual's DNA, corresponding to the gene or gene fragment. The amplified DNA is sequenced.

15 Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences, due to allelic differences. The sequences of the present invention can be used to particular advantage to obtain such identification sequences from individuals and from tissue, as further described in the Examples. The polynucleotide sequences shown in the sequence listing and the inserts contained in the deposited cDNAs uniquely represent portions of the human genome. Allelic

variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals.

5 If a panel of reagents from sequences of this invention is used to generate a unique ID database for an individual, those same reagents can later be used to identify tissue from that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

10 Another use for DNA-based identification techniques is in forensic biology. PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc. In one prior art technique, gene sequences are amplified at specific loci known to contain a large number of allelic variations, for example the DQa class II HLA gene (Erlich, H., PCR Technology, Freeman and Co. (1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene.

15 The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of the present invention. Such reagents can comprise complete genes, ESTs or corresponding coding regions, or

-26-

fragments of either of at least 20 bp, preferably at least 50 bp, most preferably at least 500 to 1,000 bp.

There is also a need for reagents capable of identifying the source of a

5 particular tissue. Such need arises, for example, in forensics when pieces of tissue of unknown origin. Appropriate reagents can comprise, for example, DNA

probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

The present application is directed to nucleic acid molecules at least 85%,

Table 1 and shown in the sequence listing or to the nucleic acid sequence referenced in

deposited cDNA, irrespective of whether they encode a polypeptide having biological activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having biological activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, for one of the uses above.

VECTORS, HOST CELLS AND PROTEIN PRODUCTION

Preferred, however, are nucleic acid molecules having sequences at least 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) or to the nucleic acid sequence of the deposited cDNA which do, in fact, encode a secreted polypeptide having biological activity. By "a polypeptide having biological activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the mature protein of the invention, as measured in a particular biological assay. "A polypeptide having biological activity" includes polypeptides that also exhibit any of the same activities as a protein of the invention in an assay in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the protein, preferably, "a polypeptide having biological activity" will exhibit substantially similar dose-dependence in a given activity as compared to the protein (i.e., the candidate

٢٧

1

polypeptide will exhibit greater activity or not more than about 35-fold less and, preferably, not more than about tenfold less activity relative to the reference protein.

-29-

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac*, *mp*, *phoA* and *lac* promoters. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, 293 and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc., *supra*; pBluescript vectors, Phagescript vectors, pNHE8A, pNHI16a, pNHI18A, pNHI46A, available from Stratagene Cloning Systems, Inc.; and pIRE9a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods

are described in many standard laboratory manuals, such as Davis *et al.*, *Basic Methods In Molecular Biology* (1986).

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to stabilize and purify proteins. For example, EP-A-0 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232262). On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. See, D. Bennett *et al.*, *J. Molecular Recognition* 8:52-58 (1995) and K. Johansson *et al.*, *J. Biol. Chem.* 270:9459-9471 (1995).

A protein of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation,

-30-

VAFAKT ARA MÜCİDİ FOLYU BEŞİNCİ

5 acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Polypeptides of the present invention include: products purified from natural

sources, including bodily fluids, tissues and cells, whether directly isolated or cultured products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed

proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

For instance, for many proteins, including the mature form(s) of a secreted protein, it is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For

instance, Ron et al., *J. Biol. Chem.*, 268:2984-2988 (1993) reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 amino-terminal amino acid residues were missing. Similarly, many examples of biologically functional C-terminal deletion mureins are known. For instance, Interferon gamma shows up to ten times higher activities by deleting 8-10 amino acid residues from the carboxy terminus of the protein (Döbeli et al., *J. Biotechnology*, 7:199-216 (1988). Furthermore, even if deletion of one or more amino acids from the N-terminus or C-terminus of a protein

25 results in modification or loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete or mature form of the protein generally will be retained when less than the majority of the residues of the complete or mature form of the protein are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal Residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

In addition to terminal deletion forms of the protein discussed above, it also

20 *Polypeptides and Fragments*

The invention further provides isolated polypeptides having an amino acid sequence encoded by a deposited cDNA, or an amino acid sequence in the sequence listing identified SEQ ID NO:Y as defined in Table 1, or a peptide or polypeptide comprising a portion of the above polypeptides. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such fragments are useful, for example, in generating antibodies against the native polypeptide.

of a polypeptide can be varied without significant effect of the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity.

Thus, the invention further includes variants of a polypeptide which show substantial biological activity or which include regions of the protein such as the portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1300-1310 (1990), wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality.

As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie, J. U. et al., *supra*, and the references cited therein. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu. substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Thus, the fragment, derivative or analog of a polypeptide shown in the figures (and sequence listing), or one encoded by the deposited cDNA, may be (i) one in

which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the above form of the polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the above form of the polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein. Thus, the mature polypeptide of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 2).

TABLE 2. CONSERVATIVE AMINO ACID SUBSTITUTIONS

Aromatic	Phenylalanine Tryptophan Tyrosine
Hydrophobic	Leucine Isoleucine Valine
Polar	Glutamine Asparagine
Basic	Arginine Lysine Histidine
Acidic	Aspartic Acid Glutamic Acid
Small	Alanine Serine Threonine Methionine Glycine

Amino acids in the protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed

5 Cunningham and Wells, *Science*

244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at

every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* or *in vivo* proliferative activity.

Of special interest are substitutions of charged amino acids with other charged 10 or neutral amino acids which may produce proteins with highly desirable improved characteristics, such as less aggregation. Aggregation may not only reduce activity but also be problematic when preparing pharmaceutical formulations, because aggregates can be immunogenic (Pinckard *et al.*, *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins

et al., *Diabetes* 36:838-845 (1987); Cleland *et al.*, *Crit. Rev. Therapeutic Drugs Carrier Systems* 10:307-377 (1993).

5 Replacement of amino acids can also change the selectivity of the binding of a ligand to cell surface receptors. For example, Ostade *et al.*, *Nature* 361:266-268 (1993) describes certain mutations resulting in selective binding of TNF- α to only one of the two known types of TNF receptors. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 227:899-904 (1992) and de Vos *et al.*, *Science* 255:306-312 (1992)).

10 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced

version of a polypeptide of the invention can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

15 Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the protein in methods which are well known in the art of protein purification.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. The polypeptides of the invention also comprise those which are at least 80% identical, 20 more preferably at least 90% or 95% identical, still more preferably at least 96%, 97%, 98% or 99% identical to a polypeptide encoded by a deposited cDNA or to the polypeptide of SEQ ID NO:Y, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

25 By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology

algorithm of Smith and Waterman (*Advances in Applied Mathematics* 2:482-489, 1981) to find the best segment of similarity between two sequences.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a polypeptide described herein is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the polypeptide of the invention. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the

reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, an amino acid sequence shown in the sequence listing or to an amino acid sequence encoded by the deposited cDNA can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

The polypeptide of the present invention could be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art.

As described in detail below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting the corresponding protein expression as described below or as agonists and antagonists capable of enhancing or inhibiting function of the protein. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" receptors of secreted proteins which are also candidate agonists and antagonists according to the present invention. The yeast two hybrid system is described in Fields and Song, *Nature* 340:245-246 (1989).

Episode Bearing Portions

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention. The epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen *et al.*, *Proc. Natl. Acad. Sci.*

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. N. and Learner, R. A. (1983) "Antibodies that react with predetermined sites on proteins," *Science*, 219:860-866. Peptides capable of eliciting protein-reactive sera are

-39-

-40-

frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals.

Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson *et al.*, *Cell* 37:767-778 (1984)

at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention.

10 The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) "General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids." *Proc. Natl. Acad. Sci. USA* 82:5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten *et al.* (1986).

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe *et al.*, *supra*; Wilson *et al.*, *supra*; Chow, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J. *et al.*, *J. Gen. Virol.* 66:2347-2354 (1985). Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen *et al.*, *supra*. Further still,

U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More

generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. *et al.* (1996) on

5 Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made

10 routinely by these methods.

Fusion Proteins

As one of skill in the art will appreciate, polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EP A 304,827; Traunecker *et al.*, *Nature* 331:84-86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric secreted protein or protein fragment alone (Fountoulakis *et al.*, *J. Biochem.* 270:3958-3964 (1995)).

Antibodies

Protein-species specific antibodies for use in the present invention can be raised against an intact protein or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

25

20

25

-41-

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl *et al.*, *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the protein of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology (Köhler *et al.*, *Nature* 256:495 (1975); Köhler *et al.*, *Eur. J. Immunol.* 6:511 (1976); Köhler *et al.*, *Eur. J. Immunol.* 6:292 (1976); Hammerling *et al.*, in: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., (1981) pp. 563-581). In general, such procedures involve immunizing an animal (preferably a mouse) with a protein antigen of the invention or, more preferably, with a protein-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56° C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting hybridoma cells

-42-

are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.* (*Gastroenterology* 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the protein antigen.

Alternatively, additional antibodies capable of binding to the protein antigen of the invention may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the protein antigen. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For *in vivo* use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. See, for review, Morrison, *Science* 229:1202 (1985); Oi *et al.*, *BioTechniques* 4:214 (1986); Cabilly *et al.*, U.S. Patent No. 4,816,567; Taniguchi *et al.*, EP 171496; Morrison *et al.*, EP 173494; Neuburger *et al.*, WO 8601533; Robinson *et al.*, WO

8702671; Bouliane *et al.*, *Nature* 312:643 (1984); Neuberger *et al.*, *Nature* 314:268 (1985).

Identification and Diagnostic Applications

Assaying protein levels in a biological sample can occur using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods (Jalkanen, M., *et al.*, *J. Cell. Biol.* 101:976-985 (1985);

Jalkanen, M., *et al.*, *J. Cell. Biol.* 105:3087-3096 (1987)). Other antibody-based

methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (^{125}I), carbon (^{14}C), sulfur (^{35}S), tritium (^3H), indium (^{113}In), and technetium (^{99m}Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying protein levels in a biological sample obtained from an individual, protein can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of protein include those detectable by X-radiography,

NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ^{131}I , ^{113}In , ^{99m}Tc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously or intraperitoneally) into the mammal to be examined for immune system disorder. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In

the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. *In vivo* tumor imaging is described in S. W.

Burchiel *et al.*, "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments" (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, S. W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982)).

Treatment of Conditions Related to Proteins of the Invention

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a protein of the invention, particularly a secreted protein, in an individual can be treated by administration of the polypeptide (in the form of a mature protein, for secreted polypeptides). Thus, the invention also provides a method of treatment of an individual in need of an increased level of the protein of the present invention comprising administering to such an individual a pharmaceutical composition comprising an amount of the isolated polypeptide of the invention effective to increase the activity level of the protein in such an individual.

Formulations

Polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of a polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g}/\text{kg}/\text{day}$ to 10 $\text{mg}/\text{kg}/\text{day}$ of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 $\text{mg}/\text{kg}/\text{day}$, and most preferably for humans between about 0.01 and 1 $\text{mg}/\text{kg}/\text{day}$ for

-45-

the hormone. If given continuously, the polypeptide is typically administered at a dose rate of about 1 $\mu\text{g}/\text{kg}/\text{hour}$ to about 50 $\mu\text{g}/\text{kg}/\text{hour}$, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the protein of the invention may be administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, drops or transdermal patch), buccaly, or as an oral or nasal spray. By "pharmaceutically acceptable carrier" is meant a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include poly(lactides) (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma- ϵ thyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly(2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res. 15*:167-277 (1981), and R. Langer, *Chem. Tech. 12*:98-105 (1982)), ethylene vinyl acetate (R. Langer et al., Id.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. (USA)* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. (USA)* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the

lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal polypeptide therapy.

For parenteral administration, in one embodiment, the polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringers solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrose; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

-46-

The polypeptide is typically formulated in such vehicles at a concentration of

Example:

about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

EXAMPLE 1. Isolation of A Selected cDNA Clone From the Deposited Sample

Each protein of the invention is related to a human complementary DNA (cDNA) clone prepared from a messenger RNA (mRNA) encoding the related protein.

THE JOURNAL OF CLIMATE

Any polyphosphate to be used for therapeutic administration must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g.,

0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution,

reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

11C in solution also provides a pharmaceutical pack or kit comprising 11C or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

vector used in construction of the cDNA library from which each cDNA clone listed in Table 1 originally was isolated. For example, where a particular clone is identified on Table 1 as being isolated in the vector "Lambda Zap," it can be seen from the following table that this cDNA clone contained in the biological deposit in Bluescript.

-49-

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
5	plasmid BA
lafmid BA	pSport1
pSport1	pCMVSPORT 2.0
pCMVSPORT 2.0	pCMVSPORT 3.0
pCMVSPORT 3.0	pCR [®] 2.1
10	

Focus 15:59- (1993). Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. See, for instance, Clark, J. M., *Nuc. Acids Res.* 16:9677-9686 (1988) and Mead, D. et al., *BioTechnology* 9: (1991).

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, each cited deposit contains at least a plasmid for each cDNA clone identified in Table 1 as sharing the same ATCC Deposit Number.

Two approaches are used herein to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1, although others are known in art. In the first, a plasmid is isolated directly by screening clones using an oligonucleotide probe. To isolate a particular clone, a specific oligonucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ³²P- γ -ATP using T4 polynucleotide kinase and purified according to routine methods (e.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY, 1982). The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor DH10B, also available from Life Technologies. See, for instance, Gruber, C. E. et al.,

15

16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., *Nucleic Acids Res.* 17:9494 (1989) and pBK (Alting-Mees, M. A. et al., *Strategies* 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both may be transformed into *E. coli* strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-.

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

-51-

WO 98/31800

-52-

Laboratory Press, pages 1.93 to 1.104), or other technique known to those of skill in the art.

An alternative approach to isolate any polynucleotide of interest in the deposited library is to prepare two oligonucleotide primers of 17-20 nucleotides derived from both ends of the determined sequence for the selected clone (i.e., within the region of SEQ ID NO:X bounded by the 5' NT of the clone and the 3' NT of the clone defined in Table 1 for each cDNA clone identified therein. These two oligonucleotide primers are used to amplify the polynucleotide of interest using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out

under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl₂, 0.01%

(w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to filter probing, clone enrichment using specific probes and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res., 21(7):1683-1684 (1993). Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcript and a primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest, is

used to PCR amplify the 5' portion of the desired full-length gene which may then be sequenced and used to generate the full length gene. This method starts with total RNA isolated from the desired source; poly A RNA may be used but is not a prerequisite for this procedure. The RNA preparation may then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RN-A ligase step. The phosphatase if used is then inactivated and the RNA is treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase. This modified RNA preparation can then be used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis-reaction can then be used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

EXAMPLE 2. Features of Proteins of the Invention

Table 1, below, describes particular features of the proteins and related nucleotide and amino acid sequences of this invention.

20

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to filter probing, clone enrichment using specific probes and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res., 21(7):1683-1684 (1993). Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcript and a primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest, is

25

FEATURES OF THE PROTEIN ENCODED BY SEQ ID NO: 1

The novel full-length chemotactic cytokine V (CCV) polypeptide exhibits significant sequence identity to a chemotactic protein isolated from the murine S100 fraction designated CP-10 (chemotactic protein, 10 kD). The chemotactic cytokine V cDNA clone contains an 1091 nucleotide insert (SEQ ID NO:1) which encodes a 103 amino acid polypeptide (SEQ ID NO:2), both shown in Figure 1. The clone was obtained from an induced endothelial cell cDNA library. A sequence alignment analysis of the deduced amino acid sequence of HEMTF185 shows that CCV shares approximately 24% identity and 69% similarity to the amino acid sequence of the murine CP-10 protein. In addition, it was determined by a BLAST analysis that the amino acid sequence of chemotactic cytokine V also exhibits approximately 31% identity and 67% similarity to the previously described rat intracellular Ca2+-binding protein. An examination of expression of chemotactic cytokine V in the HGS database reveals a widespread cell and tissue distribution of this gene. Expression of this clone was observed in a wide variety of human cDNA libraries in the Human Genome Sciences, Inc. (HGS) express sequence tag (EST) database including colon carcinoma (HCC) cell line, smooth muscle, amygdala depression, keratinocytes, uninduced endothelial cells, osteoblasts, and others.

CP-10 is a potent factor capable of extravascular recruitment of polymorphonuclear cells (PMN) and monocytes from circulation. Optimal chemotactic activity of CP-10 for murine PMN and neutrophils is in the range of 10-11 and 10-13 M, making this factor one of the most potent chemotactic factors reported to date. CP-10 is the murine homologue of a human S100 protein designated migration inhibition factor-related protein 8 (MRP8). MRP 8 can occur as a complex with an additional human S100 protein termed MRP14 (the complex has previously been reported as the cystic fibrosis antigen, calgranulin A and B, or L1 antigen). This complex can comprise as much as 10-20% of the total cytoplasmic protein content of resting neutrophils and, although a significantly lower percentage of total cytoplasmic protein content, MRP8/14 complexes can also be found in resting monocytes. There

TABLE I. FEATURES OF PROTEINS OF THE INVENTION

Protein ID (Group N)	cDNA	ATCC Deposit ID and Date												First AA	Last AA	First AA of Secreted Protein	Last AA of Secreted Protein	
		Clone ID	Clone Date	Vector	NT	NT	NT	NT	NT	NT	NT	NT	NT					
PF353-1	HEMF185	209053	05/16/97	BLUESCRIPT SK-	1	1091	1	1091	118	118	2	1		103				
PF353-2	HTXET35	209053	05/16/97	BLUESCRIPT SK-	3	887	1	887	64	64	4	1	15	16	172			
PF353-3	HT3SC28	209053	05/16/97	BLUESCRIPT SK-	5	540	1	540	19	19	6	1	22	23	88			
PF353-4	HBZAK03	209053	05/16/97	PSPON 1.0	7	520	1	520	112	112	8	1		39				
PF353-5	HDFUB43	209053	05/16/97	BLUESCRIPT SK-	9	1352	1	1352	55	55	10	1		116				
PF353-6	HEGBM39	209054	05/16/97	BLUESCRIPT SK-	11	632	1	632	88	88	12	1		150				
PF353-7	HNGBHS4	209054	05/16/97	UNI-ZAP XR	13	582	1	582	1	1	14	1		193				
PF353-8	HSALA25	209054	05/16/97	BLUESCRIPT SK-	15	1356	1	1356	115	115	16	1		324				
PF353-9	HUSAAX55	209054	05/16/97	BLUESCRIPT SK-	17	2934	1	2934	1	1	18	1		977				
PF353-10	HSXCK41	209054	05/16/97	BLUESCRIPT SK-	19	1587	1	1587	1	1	20	1	15	16	528			
PF353-11	HFKF79	209054	05/16/97	BLUESCRIPT SK-	21	1359	1	1359	1	1	22	1		452				
PF353-12	HAICCH28	209054	05/16/97	UNI-ZAP XR	23	1098	1	1098	1	1	24	1		365				

-55-

is also evidence that suggests that MRP8/14 may be released from myeloid cells, although it is not clear whether the complex is actively released as part of a response to inflammation or passively as a part of the demise of such cells during the inflammatory process.

5 The function(s) of MRP8/14 complexes, CP-10, and related S100 fraction

Ca2+-binding proteins are not entirely clear. However, it is thought that a major functional role of such proteins is in the recruitment of certain populations of immune

cells to areas of inflammation. Devery and coworkers (J. Immunol. 152, 1888-1897; 1994) have demonstrated that expression of cell surface molecules such as Mac-1,

10 which is involved in the process of cell adhesion as well as several additional cellular processes, may be influenced by prior interaction of the cell with chemoattract factors such as CP-10. These studies have also been performed *in vivo* where it was observed

that CP-10 protein accumulated on the endothelial lining of small blood vessels in LPS-inflamed footpads. Furthermore, increased levels of MRP8/14 have been observed in the sera of patients afflicted with several inflammatory diseases including

15 rheumatoid arthritis. It has also been suggested that chemoattract cytokine molecules such as CP-10 or MRP8/14 may function as a type of "calcium sink" during times of elevated intracellular levels of calcium for sustained periods of time. Alternatively, it has been suggested that MRP8/14 may function as a specific inhibitor of casein kinase

20 II activity. Although the precise functional role(s) of many of the currently defined chemotactic cytokine-like proteins containing significant regions of sequence identity to HEMFL85 are not known in any detail, a number of studies with these proteins strongly suggest one or more roles for these proteins in a variety of human disease states including rheumatoid arthritis, sarcoidosis, tuberculosis, onchocerciasis, and other chronic inflammatory disease states. As a result, the discovery of a novel

25 chemotactic cytokine-like molecule is believed to be of value in a variety therapeutic and diagnostic capacities.

Owing to the homology to CP-10 and other calcium binding proteins it is expected that the CCV polypeptide shares possess common bioactivities. The

-56-

activity of CCV may be assayed by any of several biological assays known in the art, preferably calcium binding assays. The homology to CP-10 and other calcium

binding proteins indicates that the CCV polypeptide is useful in the detection and treatment of chronic inflammatory diseases such as rheumatoid arthritis, sarcoidosis, tuberculosis and onchocerciasis.

5

FEATURES OF THE PROTEINS ENCODED BY SEQ ID NOS: 3 and 5

The full-length nucleotide sequences of two novel human cDNA clones (HTXET53 and HT3SG28) which encode splice variants of the previously reported

10 and highly related chemokines LAG-2, NKG5, and 519 have recently been identified. See for example, Hercend and Triebel (WPI Acc. No. 90-132241/17). These two clones have been designated Chemokine from Activated T-Cells-1 (CAT-1)

(HTXET53), and Chemokine from Activated T-Cells-2 (CAT-2) (HT3SG28). The HTXET53 clone was obtained from a human activated (12 hour) T-cell cDNA library and contains a 887 nucleotide insert (SEQ ID NO:3) which encodes a

15 172 amino acid polypeptide (SEQ ID NO:4), shown in Figure 2. The HT3SG28 clone was obtained from a human activated (8 hour) T-cell cDNA library and contains a 550 nucleotide insert (SEQ ID NO:5) which encodes an 88 amino acid polypeptide (SEQ ID NO:6), shown in Figure 3. The predicted amino acid sequences of the novel full-

length CAT splice variants contain several regions of nearly perfect sequence identity to the previously reported human LAG-2, NKG5, and 519 lymphokines. Alignment of the amino acid sequences shows perfect identity between the two novel molecules with LAG-2 and NKG5, with the exception of a 27 amino acid insertion near the amino terminus of HTXET53, and a 57 amino acid deletion very near the carboxy terminus of HT3SG28. The 519 amino acid sequence differs from each of the novel clones and from LAG-2 and NKG5 by an 18 amino acid deletion of the hydrophobic amino terminus.

The HTXET53 polypeptide is predicted to have a 15 amino acid secretory leader sequence. The HT3SG28 polypeptide is predicted by the computer program

-57-

PSORT to have either a 15 or a 22 amino acid leader sequence. The leader sequences are underlined in Figures 2 and 3. Applicants believe that both the shorter and longer form of the HT3SG28 polypeptides (i.e., beginning at either residue 16 or residue 23) are active.

5 Expression profiles of the two novel genes are qualitatively identical in the HGS database. Additional HGS human cDNA libraries which contain the two novel CAT clones are resting T-cells, apoptotic T-cells, activated T-cells, spleen (chronic lymphocytic leukemia), activated monocytes, pituitary, and 9 week early stage human. The mRNA expression patterns of these novel genes have not been examined by Northern blot analysis.

10 The original molecule cloned from this group the T-cell-specific clone 510. NKG5 was a term used to describe a group of identical clones isolated from a human natural killer (NK) cell cDNA library. These genes are highly related and are thought to be expressed only in NK and T-cells. A genomic clone of the gene which encodes both 510 and NKG5 consists of at least five exons and four introns which are likely responsible for the generation of the related, but unique gene products. The genomic clone also reveals a number of T-cell-specific and activation state-specific regulatory sequences indicating that expression of the gene is highly restricted to certain functions of a small subset of cell types.

15 The novel and previously described molecules discussed herein also contain approximately 33% identity with a recently reported clone designated NK-lysin. NK-lysin has been found to exhibit a potent anti-bacterial activity against such organisms as *Escherichia coli*, *Bacillus megaterium*, *Acinetobacter calcoaceticus*, and *Streptococcus pyogenes*. In addition, NK-lysin was also observed to possess a marked lytic activity against an NK-cell sensitive mouse tumor cell line (YAC-1), but had no such activity against erythrocytes. As a result, there are a number of potential therapeutic and/or diagnostic applications for a factor such as those encoded by HTXET53 and HT3SG28. Applications may include the detection and treatment of such clinical presentations as various bacterial infections, a number of lymphomas,

-58-

immunological disorders, autoimmune diseases, inflammatory diseases, various allergies, and possibly as anti-infectious agents.

FEATURES OF THE PROTEINS ENCODED BY SEQ ID NOS: 7 and 9

5 The novel Melanoma Inhibitory Activity Protein (MIA)-2 and -3 cDNA clones presented herein are shown in Figures 4 and 5. The cDNA clone HBZAK03 contains a 520 nucleotide insert (SEQ ID NO:7) which encodes a 59 amino acid polypeptide (SEQ ID NO:8), as shown in Figure 4. A BLAST analysis of the predicted amino acid sequence of HBZAK03 demonstrates that this novel clone appears to be a splice variant of another cDNA clone designated HLFBD44. The nucleotide sequence of HLFBD44 (SEQ ID NO:9) and deduced amino acid sequence (SEQ ID NO:10) are shown in Figure 5. Both of these HGS clones exhibit significant sequence identity to a human gene termed melanoma inhibitory activity (MIA) protein. BestFit analysis demonstrates that the HBZAK03 protein exhibits approximately 20% identity and 10 58% similarity to the MIA protein over a region of roughly 60 amino acids. The expression profile of the HBZAK03 cDNA in the HGS database reveals that it appears in a number of HGS human cDNA libraries in addition to the prostate cDNA library from which it was cloned. Some of the cDNA libraries in which this clone appears include fetal lung, the bone marrow cell line (RS4;11), macrophage, serum-treated smooth muscle, epileptic frontal cortex, subtracted fetal brain, HSA 172 cell line, induced endothelial cells, and others.

15 The highest sequence identity of the novel cDNA clones presented herein suggests that they may possess a function involved in the regulation of melanoma progression. The previously described MIA protein functions as a component of a highly complex and only partially characterized system of stimulatory and inhibitory factors which together dictate the progression of a melanoma. MIA is secreted by malignant melanoma cells and has the capacity to inhibit the growth of melanoma cells in culture. Investigators have examined the expression profile of the MIA gene by Northern blot and RT-PCR analysis and have determined that it is expressed in all

-59-

-60-

-60-

melanoma cell lines, a few glioma cell lines, approximately half of the benign melanomas, all malignant melanomas, and from all lymph node metastases of malignant melanomas examined (Bosserhoff et al., *J. Biol. Chem.* 271, 490-495, 1996). In contrast, no MIA expression was detected by these methods in samples obtained from any other skin-derived cells including normal fibroblasts, HaCaT keratinocytes, COS cells, HeLa cells, HepG2 cells, DU 145 (human prostate carcinoma) cells, and J82 (human bladder carcinoma) cells.

Based on the sequence similarity between these polypeptides MIA-2 and -3, are predicted to be useful in the detection and regulation of malignant melanoma, in immune system modulation, and in the treatment of cardiac arrest and stroke. Other activities of MIA-1 as well as assays for detecting MIA-1 activity are outlined in WO 95/03328, hereby incorporated herein by reference in its entirety. MIA-2 and -3 activity can be assayed accordingly.

15 FEATURES OF THE PROTEINS ENCODED BY SEQ ID NOS: 11 and 13

A macrophage-specific protein, termed AIF-1, has only very recently been molecularly cloned. AIF-1 appears to function in macrophage activation in the pathogenesis of chronic cardiac rejection following transplantation. A characteristic 20 manifestation of cardiac tissue rejection following transplantation is an immune-mediated arteriosclerosis which ultimately results in graft failure and creates the need for retransplantation during the first postoperative year. It is thought that the arteriosclerotic state results from an alloimmune response involving activated immune cells, particularly macrophages, which stimulate smooth muscle-cell migration and proliferation into the area of the transplant leading to lesions in donor vessels. AIF-1 was identified by Utans and coworkers (*J. Clin. Invest.* 95, 2954-2962, 1995) in ongoing studies of inducible gene expression patterns in macrophage cells in a chronic rejecting rat heart allograft model. AIF-1 was expressed in response to INF- γ in the chronic cardiac rejection model referenced above. Expression of AIF-1 was seen

selectively in activated macrophages, neutrophils, and the macrophage-like cell lines THP-1, U937, and HL60, but not in several other human cells and tissues examined. Furthermore, low levels of AIF-1 expression can be observed in endomyocardial biopsy samples obtained from human heart transplant patients.

5 The cDNA clone designated HEBGM49 or "AIF-2" contains a 632 nucleotide cDNA insert (SEQ ID NO:11) encoding a 150 amino acid polypeptide (SEQ ID NO:12), as shown in Figure 6. The cDNA clone was isolated from a human early stage brain cDNA library. This clone also appears in several other cDNA libraries constructed from a variety of human cell and tissue types including fetal epithelium, fetal kidney, hippocampus, tongue, and osteoblastoma HOS cells. A BLAST analysis of the amino acid sequence of HEBGM49 demonstrated that this clone exhibits approximately 65% identity and 80% similarity with AIF-1 over its entire length.

The cDNA clone HNGBH45 or "AIF-3" contains a 757 nucleotide cDNA insert (SEQ ID NO:13) encoding a 193 amino acid polypeptide (SEQ ID NO:14), as shown in Figure 7. The cDNA clone was isolated from a human neutrophil cDNA library. This clone appears in a number of additional cDNA libraries including aortic endothelium, cerebellum, corpus callosum, CD34-depleted buffy coat, activated neutrophil, colon cancer, resting T-cells, tonsils, and others. A BLAST analysis of the amino acid sequence of HNGBH45 demonstrated that this clone exhibits approximately 25% identity and 47% similarity over approximately 70 amino acids of the AIF-1 molecule.

15 AIF-2 and AIF-3 are believed to be valuable clinical markers for assessing varying degrees of acute and chronic rejection of transplanted cardiac tissue. In addition, monitoring the level of AIF-2 and/or AIF-3 expression may also be useful in determining the level of macrophage or neutrophil infiltration into area of the transplanted tissue. In addition, AIF-2 and -3 may be used as targets in assays for the identification of antagonists such as small organic molecules which act to block AIF activity. Such assays are known in the art.

-61-

FEATURES OF PROTEIN ENCODED BY SEQ ID NO: 15

The full-length nucleotide sequence of a novel human cDNA clone

(HSAA12.5) has been isolated which is believed to encode a new member of the annexin/lipocortin supergene family. The novel polypeptide is termed herein

"Annexin HSAA12.5". The annexin/lipocortin supergene family is composed of at least ten calcium-binding proteins proposed to function in a variety of cellular roles including phospholipase A2 and protein kinase C inhibition, anti-coagulation, endo- and exo-cytosis, inositol phosphate metabolism, and as calcium channel proteins.

Eukaryotic calcium-binding proteins are typically classified as proteins which bind calcium by a mechanism which either includes or does not include an E-F hand motif. The annexin/lipocortin superfamily is the largest group of calcium-binding proteins whose interaction with calcium is not mediated by an E-F hand motif. Structurally, all known annexins may be characterized by a common carboxy terminal region consisting

of four similar amino acid sequences, of approximately seventy amino acids each, termed the "annexin repeats". Conversely, the amino termini of annexin/lipocortin proteins vary widely in both length and amino acid composition between member protein sequences. Typical expression patterns of annexin/lipocortin proteins include a wide variety of cells and tissues including lung, kidney, bone marrow, spleen,

thymus, brain, macrophage, placenta, ovary, uterus, skeletal muscle, and others. Annexin/lipocortin proteins are involved in a wide variety of physiologically important cellular processes. For example, lipocortin-1 (LC-1; also known as annexin-1) appears to function as a second messenger in the anti-inflammatory glucocorticoid signal transduction cascade. Most LC-1 molecules are cell surface-associated and attached to the plasma membrane by a Ca2+-dependent interaction with unrelated plasma membrane binding molecules. The process of extravasation, in which

polymorphonuclear leukocytes (PMNs) migrate into an area of inflammation, adhere to the vascular wall, and eventually pass through the vascular wall into the surrounding tissue, may be delayed by glucocorticoids, and, as a result of LC-1

-62-

function, the overall process of inflammation may be delayed. As an example of the diversity of LC-1, and other annexin/lipocortin superfamily member, function, LC-1 has also been shown to play a major regulatory role in a number of possibly unrelated cellular systems such as cell growth regulation and differentiation, response of the CNS to cytokines, neuroendocrine secretion, anti-coagulation, and neurodegeneration.

Annexin HSAA12.5 contains a 1356 nucleotide cDNA insert (SEQ ID NO:15) encoding a 324 amino acid polypeptide (SEQ ID NO:16), as is shown in Figure 8. HSAA12.5 was isolated from a cDNA library made from the HSA 172 cell line. Although previously described annexin/lipocortin proteins are widely expressed, this clone also appears only once in the HSA 172 cell line cDNA library and does not appear in any other tissue type assayed for. A BLAST analysis of the amino acid sequence of HSAA12.5 demonstrated that this clone exhibits at least 30% identity and 53% similarity over the entire length of a molecule designated human annexin-III, a member of the annexin/lipocortin supergene family.

There is clearly a need for identifying and exploiting novel members of the annexin/lipocortin superfamily such as the cDNA clone described herein. Plasma membrane-associated molecules, such as the novel potential members of the annexin/lipocortin superfamily detailed here, should prove useful in target based screens for small molecules and other such pharmacologically valuable factors that may be useful for regulating the complex processes of inflammation. Furthermore, Annexin HSAA12.5 is believed to be useful as a regulator of coagulation (anti-coagulant) by affecting Ca2+-dependent cell to cell aggregation. In addition, this annexin-like clone may prove valuable in a number of other therapeutically useful roles as an anti-inflammatory agent including regulation of ischemia, tumor metastasis, rheumatoid arthritis, other inflammatory diseases, wound healing, arteriosclerosis, and other heart diseases.

FEATURES OF PROTEIN ENCODED BY SEQ ID NO: 17

-63-

-64-

The full-length nucleotide sequence of a novel human cDNA (HUSA/X55) which encodes a previously unidentified "ES/130-like 1" protein has been identified. The

translation product of the novel full-length ES/130-like 1 cDNA clone exhibits significant sequence identity to the chicken EDTA-soluble/130 kDa protein (ES/130) gene. The ES/130-like 1 cDNA clone contains an 3036 nucleotide insert (SEQ ID NO:17) which encodes a 977 amino acid polypeptide (SEQ ID NO:18), as shown in

Figure 9. The clone was obtained from an umbilical vein endothelial cell cDNA library. A BLAST analysis of the deduced amino acid sequence of HUSA/X55 exhibits approximately 66% identity and 83% similarity to the amino acid sequence of the chicken ES/130 gene over a 573 amino acid stretch. Expression of ES/130-like 1 is detected in a wide collection of HGS human cDNA libraries including amygdala depression, thymus, smooth muscle, endometrial tumor, synovial sarcoma, macrophage, fetal heart, and a number of others. Northern blot analyses performed on expression of the ES/130-like 1 gene indicates a high level of expression in pancreas and liver and moderate to low expression elsewhere.

The in vitro process of endothelial cell transformation to mesenchymal tissue models a similar in vivo process in the developing heart where closely associated epithelial cells undergo a transformation to cardiac mesenchyme tissue. This

transformation is a required event for the development of a multichambered heart from the primitive, single chambered heart tube. ES/130 was originally identified as a 130

kD antigen isolated from the 100,000 x g pellet fraction of non-cytolytic EDTA extracts of developing chicken cardiac tissue. Inclusion of this fraction in cardiac endothelial cell cultures results in formation of mesenchymal tissue. ES/130 is an

extracellular, secreted protein which, in addition to endothelial cell transformation, has been proposed to function in the regulation of adhesion molecule expression and limb bud ectoderm, neural tube, and notocord development. Potential therapeutic and/or diagnostic applications for the ES/130-like 1 protein include such clinical presentations as atherosclerosis, restenosis, or as a general factor following a number of types of surgery.

FEATURES OF THE PROTEIN ENCODED BY SEQ ID NO: 19

The full-length nucleotide sequence of a human cDNA clone (HSXCK41) which encodes a novel brain-enriched hyaluronan-binding factor ("BEF") has been determined. The novel BEF cDNA clone presented herein was discovered in a human substantia nigra cDNA library. The clone contains a 1757 nucleotide insert (SEQ ID NO:19) which is predicted to encode a 528 amino acid polypeptide (SEQ ID NO:20).

A BLAST analysis of the predicted amino acid sequence of HSXCK41 demonstrates significant sequence identity to the bovine brevican mRNA (GenBank entry X75887). A member of the aggrecan/versican family of cell surface proteoglycans. The HSXCK41 amino acid sequence exhibits approximately 92% identity and 93% similarity over an approximately 400 amino acid stretch of the brevican sequence. This clone has been identified in a number additional HGS human cDNA libraries, many of which originate from neural tissues. These include epileptic frontal cortex, early stage brain, skin tumor, hippocampus, cerebellum, hemangiopericytoma, infant brain, fetal brain, and fetal bone.

The aggrecan/versican family of cell surface proteoglycans may be characterized by the presence of chondroitin sulfate side chains, a hyaluronic acid (HA)-binding motif in the amino terminal domain, and at least one epidermal growth factor (EGF)-like repeat, a lectin-like motif, and one or more complement regulatory protein (CRP)-like motifs in the carboxy terminal domain. The aggrecan/versican family includes a number of members such as brevican, aggrecan, decorin, versican, and neurocan. Brevican is expressed predominantly in the brain and in primary cerebellar astrocytes, but not in neurons. Meanwhile, both aggrecan and versican are expressed in chondrocytes in human articular cartilage obtained from subjects of a wide range of ages. Aggrecan messenger RNAs undergo alternative splicing events which vary the inclusion or exclusion of the single EGF-like motif in the carboxy terminal domain. Alternatively, versican contains two EGF-like motifs and a single CRP-like motif, all of which are present in all expression patterns examined. Finally, the expression of

two recently described members of the aggrecan/versican family isolated from the

The functional roles of members of the aggrecan/versican family are rather varied. Aggrecan itself aggregates with HA to function as a major space-filling component of cartilage. Brevican, an aggrecan/versican family member which is a conditional chondroitan sulfate proteoglycan, appears in a secreted, soluble form as well as in a GPI-anchored form. Both brevican isoforms have been implicated as functional components of the terminally differentiating and adult nervous systems. It will likely be determined that molecules such as these and the novel BEF cDNA clone HSXCK41 may play a role in one or more of a variety of cellular processes which typically involve intercellular contact and communication mediated through cell surface and/or secreted glycoprotein factors. Such cellular processes might include cell adhesion, proliferation, tumor metastasis, and lymphocyte migration into areas of inflammation. Related polypeptides are believed to be expressed at a higher level in tumors such as gliomas. Thus, BEF polynucleotides and polypeptides are useful as diagnostic markers and reagents for detection of tumors such as gliomas.

THE JOURNAL OF CLIMATE

The full-length nucleotide sequence of a human cDNA clone (HFKFY79) which encodes a novel adipose differentiation factor ("ADF") has recently been determined. The novel ADF cDNA clone presented herein was originally isolated from a human fetal kidney cDNA library. The clone contains a 1550 nucleotide insert (SEQ ID NO:21) which encodes a 452 amino acid polypeptide (SEQ ID NO:22), as shown in Figure 11. A BLAST analysis of the predicted amino acid sequence of HFKFY79 demonstrates that this clone exhibits its highest degree of sequence relatedness in the GenBank public database to the murine ADF protein (GenBank accession number M92275). Based on its homology to murine ADF, human ADF is believed to share common biological activities. A BestFit analysis of the predicted amino acid sequence of HFKFY79 versus the murine ADF amino acid sequence demonstrates that the two

Protein sequences exhibit approximately 39% identity and 79% similarity. The expression profile of the HFKFY79 clone suggests a widely distributed expression pattern. In addition to the human fetal kidney library from which this clone was obtained, it also appears in a large number of human cDNA libraries including ulcerative colitis, adult testis, hypothalamus, induced endothelial cells, Jurkat T-cell line in S-phase, serum-treated and control smooth muscle, adipocytes, adult small intestine, lymph node breast cancer, infant brain, and many others.

The murine ADF gene was cloned by Jiang & Serrero (Proc. Natl. Acad. Sci. USA 89, 7856-7860; 1992, incorporated herein by reference) in an effort to identify genes whose expression profiles change significantly during the process of 1246 adipocyte cell and primary adipocyte differentiation. The murine ADF gene product identified by Jiang & Serrero is a 50 kD, membrane-bound protein expressed abundantly in mouse fat pads. The novel cDNA presented herein also exhibits sequence identity to several additional lipid-specific proteins. The first of the putative homologues is the major substrate for cAMP-dependent protein kinase A (PKA) in adipocytes and is termed perilipin. Perilipin is expressed in two alternatively spliced forms designated perilipins A and B. Both forms of perilipins are expressed exclusively at the surface of lipid storage droplets. It is thought that perilipids may function as a barrier to deny access of lipase to lipid reservoir of unstimulated cells. This event may be regulated by PKA-dependent phosphorylation of perilipin which allows exposure of lipid molecules to lipase. In addition, ADF is also related by sequence identity to a gene cloned from a human bone marrow-derived stromal cell line (KM-102) designated adipogenesis inhibitory factor (AGIF). AGIF has been shown to inhibit the process of adipogenesis in the mouse preadipocyte cell line 3T3-L1. Thus, human ADF may be useful among other things as a therapeutic modulator of lipid metabolism in the human body.

-5-

-67-

WO 98/31800

-68-

FEATURES OF THE PROTEIN ENCODED BY SEQ ID NO: 23

The novel "Bcl-like" cDNA clone (HAICH28) presented herein was originally

identified in a TNF- α /IFN-induced endothelial cell cDNA library. The clone contains
5 a 1211 nucleotide insert (SEQ ID NO:23) which encodes a 365 amino acid
polypeptide (SEQ ID NO:24). A BLAST analysis of the predicted amino acid
sequence of HAICH28 demonstrates that this clone exhibits strong sequence
similarity to two previously reported genes termed bovine polyA⁺ binding protein II
and human Bcl-w (GenBank accession numbers X89969 and U59747, respectively).10 The expression profile of the HAICH28 clone suggests a widely distributed
expression pattern. In addition to the TNF- α /IFN-induced endothelial cell library
from which this clone was obtained, it also appears in a large number of human cDNA
libraries including PHA-stimulated T-cells, osteoblasts, schizophrenic hypothalamus,
activated monocytes, adrenal gland tumor, primary dendritic cells, and a number of
others.15 The protein product of the related Bcl-w gene has been determined to function as
a key player in the cellular apoptosis or cell death pathway. Apoptosis is a term
which describes the process of programmed cell death in vertebrates. During the
process of apoptosis, the cell membrane shrinks and blebs resulting in a loss of
membrane integrity and intercellular contact. In addition, the chromatin is condensed
20 and cleaved into a characteristic ladder-like organization and, finally, vesicular
remnants of the cell are quickly engulfed and destroyed by neighboring cells. The
signal for the cell to enter the apoptotic pathway likely begins with the binding of Fas
ligand or tumor necrosis factor (TNF), or the recently discovered TRAIL ligand, to the
25 Fas/CD95/APO-1 or TNF (p55), or DR4 or DR5 receptors, respectively. These
ligand/receptor interactions recruit a cellular protein designated FLICE to the cell
membrane to act as a physical link between the Fas/CD95/APO-1 and TNF receptor
complexes, also termed death receptors, and the cysteine proteases belonging to theinterleukin-1 β (IL-1 β) converting enzyme (ICE)/CED-3 family to carry out the
process of apoptosis.The t(14;18) chromosomal translocation is often associated with human follicular
lymphoma. In this chromosomal abnormality, the immunoglobulin heavy chain locus
5 becomes translocated adjacent to the Bcl-2 gene, resulting in a drastic overexpression
of the Bcl-2 gene. Bcl-2 blocks the process of apoptosis by an unknown mechanism.
It has been proposed that Bcl-2 controls the process of apoptosis by regulating
endoplasmic reticulum-associated Ca²⁺ fluxes. Several other genes have been
identified which have significant regions of sequence identity with Bcl-2, including
10 Ced-9, BHRF1, Bax, Bcl-xS, Bcl-xL, Bcl-w, Bak, Mcl-1, and GRS. The protein
product of each of these genes can affect the process of apoptosis in either a positive
(for example, Bax or Bcl-xS) or negative (for example Bcl-2, BHRF1, Ced-9, or Bcl-
xL) fashion.A large number of cells fall victim to the apoptotic process throughout
15 development and during the lifetime of the organism. Clearly, strict regulation of the
functional molecules comprising such a potentially dangerous process is an extremely
necessary and valuable facet of the repertoire of cellular regulatory pathways. As a
result, the identification of novel molecules related to Bcl-2 or Bcl-w, such as that
encoded by the novel cDNA clone described herein, represents a major step in
20 understanding, and, in turn, exploiting the complex process of controlled cell death.
Accordingly, the Bcl-like polypeptide of the present invention is thought to be useful
as a therapeutic in an anti-viral or anti-tumor capacity or, alternatively, in a diagnostic
capacity.

25

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13(b)s)

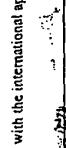
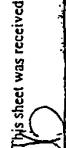
A. The indications made below relate to the microorganism referred to in the description on page <u>53</u> , line <u>N/A</u>	<input type="checkbox"/> Further deposits are identified on an additional sheet
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 1201 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit <u>May 16, 1997</u>	Accession Number <u>209053</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g. "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov	<input type="checkbox"/> This sheet was received by the International Bureau on: <u>17-7-1997</u>
<input type="checkbox"/> For receiving Office use only	<input type="checkbox"/> For International Bureau use only
Authorized officer  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov	Authorized officer  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13(b)s)

A. The indications made below relate to the microorganism referred to in the description on page <u>53</u> , line <u>N/A</u>	<input type="checkbox"/> Further deposits are identified on an additional sheet <input type="checkbox"/>
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 1201 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit <u>May 16, 1997</u>	Accession Number <u>209054</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g. "Accession Number of Deposit")	

<input checked="" type="checkbox"/> This sheet was received with the international application  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov	<input type="checkbox"/> For receiving Office use only
<input type="checkbox"/> This sheet was received by the International Bureau on: <u>17-7-1997</u>	<input type="checkbox"/> For International Bureau use only
This sheet was received by the International Bureau on:	
<input type="checkbox"/> This sheet was received with the international application  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov	
Authorized officer  Brian M. Hutchings 703-305-7453 brian.hutchings@uspto.gov	

-71-

-72-

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide selected from the group consisting of:
 - (a) the polypeptide shown in SEQ ID NO:2;
 - (b) the polypeptide shown in SEQ ID NO:4;
 - (c) the mature polypeptide shown as residues 16-172 in SEQ ID NO:4;
 - (d) the polypeptide shown in SEQ ID NO:6;
 - (e) the mature polypeptide shown as residues 16-88 in SEQ ID NO:6;
 - (f) the mature polypeptide shown as residues 23-88 in SEQ ID NO:6;
 - (g) the polypeptide shown in SEQ ID NO:8;
 - (h) the polypeptide shown in SEQ ID NO:10;
 - (i) the polypeptide shown in SEQ ID NO:12;
 - (j) the polypeptide shown in SEQ ID NO:14;
 - (k) the polypeptide shown in SEQ ID NO:16;
 - (l) the polypeptide shown in SEQ ID NO:18;
 - (m) the polypeptide shown in SEQ ID NO:20;
 - (n) the mature polypeptide shown as residues 16-528 in SEQ ID NO:20;
 - (o) the polypeptide shown in SEQ ID NO:22; and
 - (p) the polypeptide shown in SEQ ID NO:24.
2. The nucleic acid molecule of claim 1 comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, and SEQ ID NO:23.
3. An isolated nucleic acid molecule of claim 3 comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.
4. An isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule of claim 1, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.
5. An isolated nucleic acid molecule of claim 6 comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.
6. An isolated polypeptide comprising an amino acid sequence which is identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.
7. An isolated polypeptide of claim 6 comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.
8. An isolated polypeptide comprising an amino acid sequence identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.
9. A method of making a recombinant vector comprising inserting an isolated nucleic acid molecule of claim 1 into a vector.

FIG. 1A

42 L Q K E L N H M L S D T G C G N R K L R I K 61
 241 CTC CAG AAA GAG CTG AAC CAC ATG CTG TCG GAC ACA GGG AAC CCG AGG ATT CGG AGG 300
 250 260 270 280 290 300
 22 Y V S K Y S L V K N K I S K S E R E M 41
 191 TAT GTG TCT ATG TAC ACC CTG GTC AGC ACC ATC ACC AGC AGC ATG ATG 240
 190 200 210 220 230 240
 1/45 2 S D C Y T E L E K A V I L V E N F Y K 21
 121 TCA GAC TGC TAC ACC CTG GAG GAG GCA GTC ATT GTC CTG GTG CTA AAC TTC TAC AAA 180
 130 140 150 160 170 180
 61 CCT GCT GCA GAG GAG GCA GAC TGA GGC AGC AGG CCG CCG CAG CAG GCG AAC AGG GAG ATG 120
 70 80 90 100 110 120
 1 AGC CCG GCT GGG CTG AGC GCA GGG AGC TGC TGG GCA GTG CCA GAG CCC AGG CCC CAG AGC 60
 10 20 30 40 50 60
 5 vector of claim 10 into a host cell.

10. A recombinant vector produced by the method of claim 9.

11. A method of making a recombinant host cell comprising introducing a vector of claim 10 into a host cell.

12. A recombinant host cell produced by the method of claim 12.

13. A method of making an isolated polypeptide comprising culturing a recombinant host cell of claim 12 under conditions such that said polypeptide is expressed and recovering said polypeptide.

14. An isolated polypeptide produced by the method of claim 13.

15. An isolated antibody capable of specifically binding to a polypeptide of claim 6.

HEMF185 (SEQ ID NOS: 1 and 2)

310 320 330 340 350 360
 301 CTC ATC CAG AAC CTG GAT GCC AAT CAT GAT GGG CGC ATC AGC TTC GAT GAG TAC TGG ACC 360
 62 L I Q N L D A N H D G R I S F D E Y W T 81
 370 380 390 400 410 420
 361 TTG ATA GGC GGC ATC ACC GGC CCC ATC GCC AAA CTC ATC CAT GAG CAG GAG CAG CAG AGC 420
 82 L I G G I T G P I A K L I H E Q E Q Q S 101
 430 440 450 460 470 480
 421 AGC AGC TAG AGA CCC CTT TGG CCA CAC CTT CCA GGC ACT GGC CTG ATG CCC CGC CCT GGT 480
 102 S S * 103 2/45
 490 500 510 520 530 540
 481 GCT CTC CCC AGG CTC CCT CAG CCT GCC CAC CCA GGG CCC TTT ACT CTC TTC TCC 540
 550 560 570 580 590 600
 541 CTC CAG ACC TTC CTC TGA CCC TTG CTG AAC TGG GGT CCC TTT GTG AGT GTC TCA GTC TAG 600
 610 620 630 640 650 660
 601 AGG TAC CTC CCT CCC TGG GGG GTC TCA GCT CCT GGA GTC GCA GGC CCT TGG GGC CCC TCT 660

FIG. 1B

670 680 690 700 710 720
 661 GTG AGA TCT CAA TGC TGT CTG GGG ACC CTA AGA GTT TTC TCA CCT GTT CAG TCT CAT CTA 720
 730 740 750 760 770 780
 721 ACC TTC CAA TGT CTG ATG TTC CTG CCA AAT TCC TGC CTG ATT CTG GGT CCG TCC TGA CCT 780
 790 800 810 820 830 840
 781 CCA AAG GTC AGC TTG GTG CTT GAG GTC TCC CTG CTC TTG GTG GCA GTG GTA GCA GCA ACA 840
 850 860 870 880 890 900 3/45
 841 GCA GCA GCA GCA GCA GCA GCA GCA GAG ACC TCT CCA CTT TCC CTT AGC CCC TCT GCT 900
 910 920 930 940 950 960
 901 GGG TAG AGA GGC ACT TTC AGG GAC TTC CCT CCA GCT GCC TCT TCA TCT GGG AAT GAG CTA 960
 970 980 990 1000 1010 1020
 961 AGC AAG GCT GAG CCT CCT CCT GTT GCT TGA AAT AAT GAT GAT ATA AAG GCT GGA TTT GGA 1020

FIG. 1C

FIG. 2A

10	GCACCGAAGGCCCTCCCTGCCATAAGGAGGTGATGAAAGGCAATCTCAGGGCTGCC	50
70	ACCAATGGCTAACCTGGGCCCTCCCTGCCATGCAAGCCATACTCAGGGCTGCC	90
110	GAAGCTCAAGTCTGAGCCCAAGGCAAGAACACTCTCTGGAAAGGAGATGGCAATTGGCTGG	130
170	GCACATCTGGATGAGCTCTGGCTCTCTCTGGCTAGCCCTAGTACACCTGGCA	190
210	E V S V S P K G K N T S G R E S G F G W	230
250	A I W M E G L V F S R L S P E Y Y D L A	270
290	AGAGGCCACCTGGCTGAGATGAGAATCTCCCTGCCCGTGCCTGGCCACAGAGGAGGCC	310
350	R A H L R D E E K S C P C L A Q E G P Q	330
350	GGTGAACCTGGACCAAAACAGGAACTGGCCGCTGACCTACAGGACCTGGTCTGACGATA	310

HTXET53 (SEQ ID NOS: 3 and 4)

FIG. 1D

430	450	470
ACCCGGGTGTGTAAGCACGGGAGGTACGATGGCGCGACGTCTGCAGAAATTTCATGAGG		
T R V C R T G R S R W R D V C R N F M R		
490	510	530
AGGTATCAGTCTAGAGTTACCCAGGGCCTCGTGGCCGGAGAAACTGCCCAGCAGATCTGT		
R Y Q S R V T Q G L V A G E T A Q Q I C		
550	570	590
GAGGACCTCAGGTTGTATACCTTCTACAGGTCCCCCTGAGCCCTCTCACCTTGTCCCT		
E D L R L C I P S T G P L *		
610	630	650
GTGGAAGAACAGCACAGGCTCCTGTCCTCAGATCCGGGAAACCTCAGCAACCTCTGCCGGCT		
670	690	710
CCTCGCTTCCTCGATCCAGAATCCACTCTCCAGTCTCCCTCCCCCTGACTCCCTCTGCTGT		
730	750	770
CCTCCCTCTCACCGAGAATAAGTGTCAAGCAAGATTAGCCGCAGCTGCTCTTCTTT		
790	810	830
GGTGGATTGAGGGGTGGGTGTCAGTGGCATGCTGGGTGAGCTGTGTAGTCCTCAATA		
850	870	
AATGTCGTGTCGTGTGTCCTCATAAAAA		

Ch 19

100

16/1/2018/00000

- 100815/86 000

PCR/0098/00960

FIG. 2B

T3SG28 (SEQ ID NOS:5 and 6)

10	30	50
CTCAGCGGCTGCCACCATGGCTACCTGGGCCCTCCTGCTCCTTGAGCCATGCTCCTG		
M A T W A L L L L A A M L L		
70	90	110
GGCAACCCAGGTCTGGTCTTCTCTCGTCTGAGCCCTGAGTACTACGACCTGGCAAGAGCC		
G N P G L V F S R L S P E Y Y D L A R A		
130	150	170
CACCTGCGTGTGAGGAGAAATCCTGCCCGTGCCTGGCCCAAGGAGGGCCCCCAGGGTGAC		
H L R D E E K S C P C L A Q E G P Q G D		
190	210	230
CTGTTGACCAAAACACAGGAGCTGGGCCGTGACTACAGGACCTGTCTGACGATAAGTCCAA		
L L T K T Q E L G R D Y R T C L T I V Q		
250	270	290
AAACTGAAGAAGATGGTGGATAAGCCCACCCCAAGGTCCCCCTGTGAGCCCTCTCACCTGT		
K L K K M V D K P T P G P L *		

7/45

FIG. 3A

SUBSTITUTE SHEET (RULE 26)

BNSDOCID: WO 98318000A3 | >

SUBSTITUTE SHEET (RULE 26)

BNNSDOCID: <WO_9831800A2-1>

(SEQ ID NOS: 7 and 8)
 CGA CCA CCG GTC CCG GTT CGC C1
 AGC TGG GAG CAG CCG TGC CTC C2
 T V V A V G L 130
 190
 CAG GC CAA TGA GCA GAT ATG GAG C
 Q A M K H M E P
 250
 GCC TTCA GAT GGT GCG CTAT TATA
 AT CGA CCG GAA TTA GAT GCT TTA A
 310
 A F S G Y Y R
 370
 A A A T C A A T G A A G C T A A A G A T T
 T G A A T T A G T C G T A T T A G T
 430
 490
 C C A G A G C T A C A A T T T A C A A A A

F-16.3B

00813180

PC-TAIPEI 09600/8651

卷之三

9/45

Hbzak03 (SEQ ID NOS: 7 and 8)

HLFBD44 (SEQ ID NOS:9 and 10)

10/45

10	30	50
CTGAGCTGGATGAGCCGTGCTCCGGTGGAAAGCAAGGGAGCCCAGCCGGAGCCATGGCC		
		M A
70	90	110
AGTACAGTGGTAGCAGTTGGACTGACCATTGCTGCTGCAGGATTGCAAGGCCGTTACGTT		
S T V V A V G L T I A A A G F A G R Y V		
130	150	170
TTGCAAGCCATGAAGCATATGGAGCCTCAAGTAAAACAAGTTTCAAAGCCTACCAAAA		
L Q A M K H M E P Q V K Q V F Q S L P K		
190	210	230
TCTGCCTTCAGTGGTGGCTATTATAGAGGTGGGTTGAACCCAAAATGACAAAACGGGAA		
S A F S G G Y Y R G G F E P K M T K R E		
250	270	290
GCAGCATTAACTAGGTGTAAGCCCTACTGCCAATAAAGGGAAAATAAGAGATGCTCAT		
A A L I L G V S P T A N K G K I R D A H		
310	330	350
CGACGAATTATGCTTTAAATCATCCTGACAAAGGAGGATCTCCTTATATAGCAGCCAAA		
R R I M L L N H P D K G G S P Y I A A K		
370	390	410
ATCAATGAAGCTAAAGATTACTAGAAGGTCAAGCTAAAAAATGAAGTAAATGTATGATG		
I N E A K D L L E G Q A K K *		

FIG.5A

11/45

430	450	470
AATTTAAGTCGATTAGTTATGTTATGAGTACTAAGTTTATAATAAAATGCCTC		
490	510	530
AGAGCTACAATTTAAAAATGATTTAGCACAAGCTAAATCTCAAAGCCTGGTATAATT		
550	570	590
TTCTTGGTTAAATTGGGGATTTAAATCAGATTATAGTTAGAATATTGCGTATTAAT		
610	630	650
TATGGCAAGCACACACCTCTGAATAGAAATTGTTCAATTACTCATTTAGCAGATAAT		
670	690	710
TTGGGACCTATGTCTACTTTCAAGGCAAAGTGAAGATGACAGTCCTGCTCTCAGGGAG		
730	750	770
CCCCCCTTAATGGGAGACTGATAAACTGGTAATTAGACTGTGATAAAAGTATGATGG		
790	810	830
AAATTAGCTTAAGCTTTAAGTAGGGACTCTCTTATTGGTGGAAAGGCTGTTCCAGG		
850	870	890
TACAGGCAACTGGCTGGCAACTGGATACTGGAACCTGTATTAAAAGTGAATTAA		
910	930	950
CCACAACTGAGACCTAAGAAATTGACCTAGGGGTGTGTGTGTGTATTCTATGTACA		
970	990	1010
TATAAACCCATTTTATTCATGCATTAAAATAGTATGATAAAGATTTCAGAGTACAGG		
1030	1050	1070
TCTGGTACAATCACAGTTCAATTGCAGCCTCAACCTCCTGGGTTAACGAGTCCTCCGCC		
1090	1110	1130
TCAGCCTCCAAAAGTACTGGGATTACAGGCATGAGTATTACATTGTATTCAAGCTAGCCC		

FIG.5B

FIG. 6A

13/45

WO 98/31800

WO 98/31800

PCT/US98/00960

I S E V T G G V S D T I S Y R D F V N M
 ATCTCAAGGTTGACAGGAGGGTCAAGTGCACATACTACCGAGACATTGGTGAACATG
 370 390 410
 L K R M M E K L G V P K T H L E M K K M
 TTAAGAGGATGAGATGAGAAGCTTGCTGCACCCAGACCCCTGGAGAATGAGAGATGAG
 310 330 350
 TTCAAGAGGATGAGATGAGAAGCTTGCTGCACCCAGACCCCTGGAGAATGAGATGCT
 250 270 290
 GAGTTCTGACAGGAGTACAGTGCAGAGAGACCTCCAGAACGATCAGGCC
 190 210 230
 GGAAAGCCGCTTCGGCCTCTCAAAAGCCGGCAAGAGGAGCTGGCCGAGATCACCGA
 130 150 170
 GAGCCGGAGCCGGACCCCTGGCTCCCTCTCCCTGCGCCGGACCT
 70 90 110
 CACGAGGCCGGACACGCCCTGGCTCCCTCTCCCTGCGCCGGACCT
 10 30 50

HEBM49 (SEQ ID NOS: 11 and 12)

FIG. 5C

12/45

WO 98/31800

TGATGTTAAGGCACAAAAAAAGAAAAAA
 1330 1350
 TTCTCAATCACATATTAAATTGGACCTTAAATGAAATTACATATAATTAA
 1270 1290 1310
 ACTCTGCAATTGTTAGTAACTTAAATTCTCCAGTGGCTATTGACATAATG
 1210 1230 1250
 CTTAAAGGTTGACACCATTATAATTCTCCAGTGGCTATTGACATAATCA
 1150 1170 1190

14/45

430 450 470

ATGCTGGGAAACGGTCGGCTGTCCTCAAGTTAGTCATGATGTTGAAGGAAAAGCCAAC
 M L G K R S A V L K L V M M F E G K A N
 490 510 530

GAGAGCAGCCCCAAGCCAGTTGGCCCCCTCCAGAGAGAGACATTGCTAGCCTGCCCTGA
 E S S P K P V G P P P E R D I A S L P *
 550 570 590

GGACCCCGCCCTGGACTCCCCAGCCTTCCCACCCCATAACCTCCCTCCCGATCTGCTGCC
 610 630

TTCTTGACACACTGTGATCCGGCACGAGCGGC

FIG. 6B

HNGBH54 (SEQ ID NOS:13 and 14)

10 20 30 40 50 60

1 ATG GGC AGC GCG GAC TGC GAG CTG AGC GCC AAG CTG CTG CGG CGC GCA GAC CTC AAC CAG 60
 1 M G S A D C E L S A K L L R R A D L N Q 20

70 80 90 100 110 120

61 GGC ATC GGC GAG CCC CAG TCG CCC AGC CGC CGC GTC TTC AAC CCC TAC ACC GAG TTC AAG 120
 21 G I G E P Q S P S R R V F N P Y T E F K 40

130 140 150 160 170 180

121 GAG TTC TCC AGG AAG CAG ATC AAG GAC ATG GAG AAG ATG TTC AAG CAG TAT GAT GCC GGG 180
 41 E F S R K Q I K D M E K M F K Q Y D A G 60

190 200 210 220 230 240

181 CGG GAC GGC TTC ATC GAC CTG ATG GAG CTA AAA CTC ATG ATG GAG AAA CTT GGG GCC CCT 240
 61 R D G F I D L M E L K L M M E K L G A P 80

250 260 270 280 290 300

241 CAG ACC CAC CTG GGC CTG AAA AAC ATG ATC AAG GAG GTG GAT GAG GAC TTT GAC AGC AAG 300
 81 Q T H L G L K N M I K E V D E O F D S K 100

15/45

FIG. 7A

FIG. 8A

10	30	50	ACATAATTACATTGAGATTAAACAGTGAACCTTAAATTCTTCTGCTTCACTAGAACACA
70	90	110	GTTTATGCATACTCAAATAATTCTACCTCTGAGCTTAACTTACATACTCAAATAATTGTT
130	150	170	TGTGAGACCTAGTGCACAAAGAACACCATCTTCCAGCTCCAAATTCTCCATAATTGAT
190	210	230	C G D Y V Q O G T I F P A P N F N P I M D
250	270	290	GCCCAAGATGCTAGGAGAGCACCTCCAGCTTACAGGATTTGACATGCAAGACATGCTACAC
310	330	350	A Q M L G G A L Q G F D C D K D M L I N
370	390	410	ATTCCTGACTCTCAGCTGCAATGCAATGAGCATGATGAGCTTACAGAGCATG
370	390	410	TATGCCGCGAACCTGAGATTGGGGATCTGAGGAGCAGCTTCTGGATACACTCAAGAGATG
370	390	410	Y G R D L I G D L R E Q L S D H F K D V
370	390	410	ATGCCCTGACCTGACACATGCACTGCTGAGCTTACAGGCTTCAGCATG
370	390	410	M A G L M Y P P L Y D A H E L W H A M

HSAAL25 (SEQ ID NOS:15 and 16)

310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	582	5841	5861	5881	5901	5921	5941	5961	5981	5991																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
301	CTG	AGC	TTC	CGG	GAG	TTC	CTC	CTG	ATC	TTC	CGC	AGG	GCG	GCG	GCG	GAG	CTT	CGG	GAG	ATG	GAG	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	582	5841	5861	5881	5901	5921	5941	5961	5981	5991																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
501	L	S	F	R	E	L	I	F	R	K	A	A	G	E	L	Q	E	E	L	Q	E	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	10010	10011	10012	10013	10014	10015	10016	10017	10018	10019	10020	10021	10022	10023	10024	10025	10026	10027	10028	10029	10030	10031	10032	10033	10034	10035	10036	10037	10038	10039	10040	10041	10042	10043	10044	10045	10046	10047	10048	10049	10050	10051	10052	10053	10054	10055	10056	10057	10058	10059	10060	10061	10062	10063	10064	10065	10066	10067	10068	10069	10070	10071	10072	10073	10074	10075	10076	10077	10078	10079	10080	10081	10082	10083	10084	10085	10086	10087	10088	10089	10090	10091	10092	10093	10094	10095	10096	10097	10098	10099	100100	100101	100102	100103	100104	100105	100106	100107	100108	100109	100110	100111	100112	100113	100114	100115	100116	100117	100118	100119	100120	100121	100122	100123	100124	100125	100126	100127	100128	100129	100130	100131	100132	100133	100134	100135	100136	100137	100138	100139	100140	100141	100142	100143	100144	100145	100146	100147	100148	100149	100150	100151	100152	100153	100154	100155	100156	100157	100158	100159	100160	100161	100162	100163	100164	100165	100166	100167	100168	100169	100170	100171	100172	100173	100174	100175	100176	100177	100178	100179	100180	100181	100182	100183	100184	100185	100186	100187	100188	100189	100190	100191	100192	100193	100194	100195	100196	100197	100198	100199	100200	100201	100202	100203	100204	100205	100206	100207	100208	100209	100210	100211	100212	100213	100214	100215	100216	100217	100218	100219	100220	100221	100222	100223	100224	100225	100226	100227	100228	100229	100230	100231	100232	100233	100234	100235	100236	100237	100238	100239	100240	100241	100242	100243	100244	100245	100246	100247	100248	100249	100250	100251	100252	100253	100254	100255	100256	100257	100258	100259	100260	100261	100262	100263	100264	100265	100266	100267	100268	100269	100270	100271	100272	100273	100274	100275	100276	100277	100278	100279	100280	100281	100282	100283	100284	100285	100286	100287	100288	100289	100290	100291	100292	100293	100294	100295	100296	100297	100298	100299	100300	100301	100302	100303	100304	100305	100306	100307	100308	100309	100310	100311	100312	100313	100314	100315	100316	100317	100318	100319	100320	100321	100322	100323	100324	100325	100326	100327	100328	100329	100330	100331	100332	100333	100334	100335	100336	100337	100338	100339	100340	100341	100342	100343	100344	100345	100346	100347	100348	100349	100350	100351	100352	100353	100354	100355	100356	100357	100358	100359	100360	100361	100362	100363	100364	100365	100366	100367	10036

WO 98/31800

18/45

PCT/US98/00960

430 450 470
 AAGGGAGTAGGCAGTGAGAATTGCCTCATTGAAATACTAGCTCAAGAACAAATGGA
 K G V G T D E N C L I E I L A S R T N G
 490 510 530
 GAAATTTCCAGATGCGAGAAGCCTACTGCTTGAATACAGCAATAACCTCCAAGAGGAC
 E I F Q M R E A Y C L Q Y S N N L Q E D
 550 570 590
 ATTATTCAGAGACCTCGGGACACTTCAGAGATACTCTCATGAACCTGGTCCAGGGGACC
 I Y S E T S G H F R D T L M N L V Q G T
 610 630 650
 AGAGAGGAAGGATATACAGACCCCTGCGATGGCTGCTCAGGATGCAATGGTCTATGGAA
 R E E G Y T D P A M A A Q D A M V L W E
 670 690 710
 GCCTGTCAGCAGAAGACGGGGAGCACAAACCATGCTGCAAATGATCCTGTGCAACAAG
 A C Q Q K T G E H K T M L Q M I L C N K
 730 750 770
 AGCTACCAGCAGCTGCGGCTGGTTTCCAGGAATTCAAAATATTCTGGCAAGATATG
 S Y Q Q L R L V F Q E F Q N I S G Q D M
 790 810 830
 GTAGATGCCATTAATGAATGTTATGATGGATACTTCAGGAGCTGCTGGTTGCAATTGTT
 V D A I N E C Y D G Y F Q E L L V A I V
 850 870 890
 CTCTGTGTTCGAGACAAACCAGCCTATTTGCTTATAGATTATAGTGCATTGAC
 L C V R D K P A Y F A Y R L Y S A I H D

FIG. 8B

WO 98/31800

19/45

PCT/US98/00960

910 930 950
 TTTGGTTCCATAATAAAACTGTAATCAGGATTCTCATTGCCAGAAGTGAATAGACCTG
 F G F H N K T V I R I L I A R S E I D L
 970 990 1010
 CTGACCATAAGGAAACGATACAAAGAGCGATATGGAAAATCCCTATTCATGATATCAGA
 L T I R K R Y K E R Y G K S L F H D I R
 1030 1050 1070
 AATTTGCTTCAGGGCATTATAAGAAAGCACTGCTGCCATCTGTGCTGGTGTGAG
 N F A S G H Y K K A L L A I C A G D A E
 1090 1110 1130
 GACTACTAAATGAAGAGGACTTGGAGTACTGTGCACTCCTCTTCTAGACACTTCCAAA
 D Y *
 1150 1170 1190
 TAGAGATTTCTCACAAATTGTACTGTTCATGGCACTATTAACAAACTATACAATCAT
 1210 1230 1250
 ATTTTCTCTTCTATCTTGAAATTATTCTAACGCCAAGAAAATATGAATGAAAGTATAT
 1270 1290 1310
 GATACTGAATTGCTACTATCCTGAATTGCTACTATCTAACGCCAAGAAAATATGAATGAAAGTATAT
 1330 1350
 TGTGCATGATGGAATAATAAAAAAAAAAAAAAA

FIG. 8C

FIG.9B

21/45

141 K K E K V A K V E P A V S S V V N S I 160
 421 AGAGAGAGAAAAAGTGCACAAAGTGCACAGCTGAGCTCTGATTCATC 480

430 440 450 460 470 480

121 V A T V P A M P Q E K L A S S P K D K K 140
 361 GTCGCCACAGTTCACAGCCATGCGCCAGAGCTGGCCTCCCTCCACAAAGAACAG 420

370 380 390 400 410 420

101 R A P A V A V A P T P V Q P I I V A P 120
 301 CGGCCTCTGCTGCTGGCTGCCTCCACAGCTGCACCCACATTATCGTCTCT 360

310 320 330 340 350 360

81 G K I P D H D P A N V T V L L R E P V 100
 241 GGAGATACCTGATCATGAGCCACATGACTGCCTCCATGAGACATG 300

250 260 270 280 290 300

FIG.9A

20/45

61 K K E K T V E K K T K K E E K P N 80
 181 AGAGAGAGAACAGTGGAGAACAGAACAGAACAGAACAGAACCTAAT 240

190 200 210 220 230 240

41 A L A N Q R K E M A K T H Q K V E K K 60
 121 GGCCTAGCACACAGCGAACAGGAGATGGCGAACATCACACAGAACAGAAC 180

130 140 150 160 170 180

21 S A I G I F L V S T F S M K E T S Y E E 40
 61 TCTGCCATTGGCATCTTCTGGACTTCCATGAGAACGCTCATGAGAACAGAAC 120

70 80 90 100 110 120

1 M D I Y D T Q T L G V V F G F M V V 20
 1 ATGGATATTACGACACCTCAACCTGGGGTTGGCTCTTGAGAGATCTGAGTGT 60

22/45

	490	500	510	520	530	540
481	CAGGTTCTCACTTCGAAGGCTGCCATCTGGAAACTGCTCCCAAGGAGGGCAGAAATACA					540
161	Q V L T S K A A I L E T A P K E G R N T					180
	550	560	570	580	590	600
541	GATGTGGCCCAGAGCCCAGAGGCACCAAAGCAAGAGGCTCCTGCCAAGAAGAAGTCTGGT					600
181	D V A Q S P E A P K Q E A P A K K K S G					200
	610	620	630	640	650	660
601	TCAAAGAAAAAGGGCCCCAGATGCCGACGGCCCTCTCACCTCCCTACAAGACGCTG					660
201	S K K K G P P D A D G P L Y L P Y K T L					220
	670	680	690	700	710	720
661	GTCTCACGGTTGGGAGCATGGTGTCAACGAGGGCGAGGCCAGCGGCTCATCGAGATC					720
221	V S T V G S M V F N E G E A Q R L I E I					240

FIG.9C

23/45

	730	740	750	760	770	780
721	CTGTCTGAGAAGGCTGGCATTCAGGACACCTGGCACAGGCCACTCAGAAGGGTGAC					780
241	L S E K A G I I Q D T W H K A T Q K G D					260
	790	800	810	820	830	840
781	CCTGTGGCGATTCTGAAACGCCAGCTGGAAGAGAAGGAAAAACTGCTGCCACAGAACAG					840
261	P V A I L K R Q L E E K E K L L A T E Q					280
	850	860	870	880	890	900
841	GAAGATGCGGCTGTCGCCAAGAGCAAATGAGGGAGCTAACAGGAGATGGCAGCAGAA					900
281	E D A A V A K S K L R E L N K E M A A E					300
	910	920	930	940	950	960
901	AAGGCCAAAGCAGCAGCCGGGAGGCCAAAGTGAAGAAGCAGCTGGTGGCCGGAGCAG					960
301	K A K A A A G E A K V K K Q L V A R E Q					320

FIG.9D

FIG. 9F

481 Q M A E L H S K L Q S S E A E V R S K C 500
 1441 CAGATGCCAGCTGACAGCAAGTACAGTCCGAGGGAGGCAAGAACATGC 1500
 1450 1460 1470 1480 1490 1500
 461 T Q S S H A S L R A D A E K A Q E Q Q Q 480
 1381 ACGCAAGCACAGCACAGCCAGCTCCGGAGAAGGCAGCCAGCACAG 1440
 25/45 1390 1400 1410 1420 1430 1440
 441 R E S E A L Q K R L D E V S R E L C H 460
 1321 AGGAGAGTGAAGGAGGCCCTGCAAGAGGCCCTGAGACAGCTAGCGAGCTGTGCAAC 1380
 1330 1340 1350 1360 1370 1380
 421 A L E A K A A F E K Q V L Q L Q A S H 440
 1261 GCTCTGGAGAGCCAGCACAGCTCCAGAGCTCTGCAAGCTGCAAGCTGCCAC 1320
 1270 1280 1290 1300 1310 1320
 WO 9813181...

FIG. 9E

401 V S K E L V E K S E A V R Q D E Q Q R K 420
 1201 GTCAAGCAAGAGCTGGAGAGTCAGAGCTGCTGGCAAGATGAGCGAGAA 1260
 1210 1220 1230 1240 1250 1260
 381 S Q V E S K Q N A E L A K L R Q E L S K 400
 1141 AGCCAGTGGAGAGCAAGCACAGCAAGCTGGCCAGCTCCAGAGCTCACAG 1200
 1150 1160 1170 1180 1190 1200
 361 Q L A R L Q O E N S I L R D A L N Q A T 380
 1081 CAGCTGGCCCGCTGCAAGAGAACATCCCTGGAGATGCTGACAGCCACG 1140
 24/45 1090 1100 1110 1120 1130 1140
 341 Q L Q G K I R T L Q E Q L E N G P N T 360
 1021 CAGCACCTGCAAGGCCAGATCCCGAACCTTCAGAGCACTGAGAATGGCCACACG 1080
 1030 1040 1050 1060 1070 1080
 321 E I T A V Q A R M Q A S Y R E H V K E V 340
 961 GAGATCACCGCTGTGGCAAGCCAGCATGCAAGCCATCGAACGCTGAGAAGAGTG 1020
 970 980 990 1000 1010 1020
 WO 9813181...

1510 1520 1530 1540 1550 1560
 1501 GAGGAGCTGAGTGGCCTCCACGGGCAGCTCCAGGAGGCCAGGGCAGAGAACTCCCAGCTC 1560
 501 E E L S G L H G Q L Q E A R A E N S Q L 520

1570 1580 1590 1600 1610 1620
 1561 ACAGAGAGAATCCGTTCCATTGAGGCCCTGCTGGAGGCCAGGCAGCGCGGGATGCCAG 1620
 521 T E R I R S I E A L L E A G Q A R D A Q 540

1630 1640 1650 1660 1670 1680
 1621 GACGTCCAGGCCAGCCAGGCAGGCTGACCAAGCAGCAGACTCGCCTCAAGGAGCTGGAG 1680
 541 D V Q A S Q A E A D Q Q Q T R L K E L E 560

1690 1700 1710 1720 1730 1740
 1681 TCCCAGGTGTGGGTCTGGAGAAGGAGGCCATCGAGCTCAGGGAGGCCGTCAGCAGCAG 1740
 561 S Q V S G L E K E A I E L R E A V E Q Q 580

1750 1760 1770 1780 1790 1800
 1741 AAAGTGAAGAACAAATGACCTCCGGGAGAAGAACTGGAAGGCCATGGAGGCAGTGGCCACCG 1800
 581 K V K N N D L R E K N W K A M E A L A T 600

FIG.9G

1810 1820 1830 1840 1850 1860
 1801 GCCGAGCAGGCCCTGCAAGGAGAAGCTGCACTCCCTGACCCAGGCCAGGAGGAATCGGAG 1860
 601 A E Q A C K E K L H S L T Q A K E E S E 620

1870 1880 1890 1900 1910 1920
 1861 AAGCAGCTCTGTCTGATTGAGGCCAGACCATGGAGGCCCTGCTGGCTCTGCTCCCAGAA 1920
 621 K Q L C L I E A Q T M E A L L A L L P E 640

1930 1940 1950 1960 1970 1980
 1921 CTCTCTGTCTGGCACAAAGAATTACACCGAGTGGCTGCAGGATCTAAAGAGAAAGGC 1980
 641 L S V L A Q Q N Y T E W L Q D L K E K G 660

1990 2000 2010 2020 2030 2040
 1981 CCCACGCTGCTGAAGCACCCGCCAGCTCCCGCGGAGGCCCTCCTCGGACCTGGCCTCCAAG 2040
 661 P T L L K H P P A P A E P S S D L A S K 680

2050 2060 2070 2080 2090 2100
 2041 TTGAGGGAGGCCAGGAGACGCAGAGCACACTGCAGGCCAGTGTGACCAAGTACCGCAGC 2100
 681 L R E A E E T Q S T L Q A E C D Q Y R S 700

FIG.9H

FIG. 9J

2530 2540 2550 2560 2570 2580
 2470 2480 2490 2500 2510 2520
 2410 2420 2430 2440 2450 2460
 2401 CAAATCAGCTCGAGATGCCGCCAACAGGAAAGCCAGATGCTGGCCCTG 2460
 801 Q S Q L D A A K S E A Q K Q S D E L A L 820
 2461 GTCAAGGCCAGCTTGAAATGAGAGCCACATTGAGATGGCTGACATGGGCC 2520
 821 V R Q Q L S E M K S H V E D G I A G A 840
 2531 CCAAGCTTCCCTCCAGAGGCCGCCAGCCAGCAAGCACCCTGAGACGGCAAG 2580
 24/45 841 P A S S P E A P P A E Q D P V Q L K T Q 860
 2581 CTGAGTGGACAGAGCCATCTGGAGATGAGACAGACAGCGAGCACTGAG 2640
 861 L E W T E A I L E D E Q T Q R Q K L M A 880
 2641 GAGTTGAGAGGCTCACAGACCTCGGCACTGAGATGGAGAAATTGAGAGCTCCGC 2700
 881 E F E A Q T S A C R L Q E L E K L R 900

FIG. 9I

781 A E C Q N Y A K E V A G L R Q L L E S 800
 2341 GCCGAGTGCACAGAACATACGCCAACAGGCTGAGGCCACCTCTCTAGAACATCT 2400
 2350 2360 2370 2380 2390 2400
 761 V R E H T S H L E A E L K H M A A S 780
 2281 GTAGAGGCCACACGCTCGCAATTGGAGGCAAGCTGGAAAGAACATGGCCGCCAAC 2340
 2290 2300 2310 2320 2330 2340
 741 V K H L E E I V E K L K G E L E S S D Q 760
 2221 GTAGAGCATCTCGAGAGATTTGAGAGCTAAAGAGAACATGGAGACCCAG 2280
 2230 2240 2250 2260 2270 2280
 721 Q V W R A K V G A A E E B L Q K S R V T 740
 2161 CAGCTGTGGAGGCCAACAGGCTGGGCCGCCAGAACAGGAGCTCCAGAACATGGCTGCCAAC 2220
 2170 2180 2190 2200 2210 2220
 701 I L A E T E G M L R D L Q K S V E E E 720
 2101 ATCCCTGGAGAGCCAGCTCACAGAGCTGCAAGAGAGCTGGAGAGAG 2160
 2110 2120 2130 2140 2150 2160

2710 2720 2730 2740 2750 2760
 2701 ACAGCCGGCCCCCTAGAGTCTTCAGAAACAGAGGAGGCTCACAGCTGAAGGAGAGACTA 2760
 901 T A G P L E S S S E T E E A S Q L K E R L 920
 2770 2780 2790 2800 2810 2820
 2761 GAAAAAGAGAAGAAGTTAACAAAGTGACCTGGGGCGGCCACGAGACTGCAGGAGCTT 2820
 921 E K E K K L T S D L G R A A T R L Q E L 940
 2830 2840 2850 2860 2870 2880
 2821 CTGAAGACGACCCAGGAGCAGCTGGCAAGGGAGAAGGACACGGTGAAGAAGCTGCAGGAA 2880
 941 L K T T Q E Q L A R E K D T V K K L Q E 960
 2890 2900 2910 2920 2930
 2881 CAGCTGGAAAAGGCAGAGGACGGCAGCAGCTCAAAGGAGGGCACCTCTGTCTGA 2934
 961 Q L E K A E D G S S S K E G T S V * 977

30/45

FIG.9K

HSXCK41 (SEQ ID NOS:19 and 20)

10 20 30 40 50 60
 1 ATGGCCCAGCTGTTCTGCCCTGCTGGCAGCCCTGGTCTGGCCAGGCTCCTGCAGCT 60
 1 M A O L F L P L L A A L V L A Q A P A A 20
 70 80 90 100 110 120
 61 TTAGCAGATGTTCTGGAAGGAGACAGCTCAGAGGACCGCGCTTTCGCGTGCATCGCG 120
 21 L A D V L E G D S S E D R A F R V R I A 40
 130 140 150 160 170 180
 121 GGCGACGCGCCACTGCAGGGCGTGGCTGGCGGCCCTCACCATCCCTGCCACGTCCAC 180
 41 G D A P L Q G V L G G A L T I P C H V H 60
 190 200 210 220 230 240
 181 TACCTGCGGCCACCGCCGAGCCGCCGGCTGTGCTGGCTCTCCGGGTCAAGTGGACT 240
 61 Y L R P P R S R R A V L G S P R V K W T 80

31/45

FIG.10A

FIG. 10C

541	GTTGACCCGGATAGAACCTCTATGATCTGATGCTACCTGAAAGACCTGATAATGAGAACCTG	550	560	570	580	590	600	181	V D P D L Y D V C Y A E D L N G E L	200
601	TTCCTGGTACCCCTCAAGAGAACCTGAACTGGACATAGAACCTGAAAGACCTGAACTGGAG	610	620	630	640	650	660	201	F L G D P P E K L T L E A R A Y C Q E	220
660	CGGGTGCAGAGATTCGCAACCGGCAACACTGATGCAAGCCCTGGAGATGGCCCTGGAGC	670	680	690	700	710	720	221	R G A E I A T T G Q L Y A W D G G L D	240
720	CGGGTGCAGAGATTCGCAACCGGCAACACTGATGCAAGCCCTGGAGATGGCCCTGGAGC	730	740	750	760	770	780	221	R G A E I A T T G Q L Y A W D G G L D	240
780	CACTGCAGCCAGGCTAGCTGATGCAAGTGCCTGCGCTACCCATCGTACACCAAGC	790	800	810	820	830	840	241	H C S P G W L A D G S V R Y P I V T P S	260

FIG. 10B

32/45	250	260	270	280	290	300	81	TTCCTGTCGCCGGAGGCAAGGCTGCTGGCGCGGGAGTGGCTGCGCTAGTG 300	F L S R G R E A E V L V A R G V R K V 100
	310	320	330	340	350	360	101	AACGAGGCCATACCGGTTCCGGCATCTGGCTCGCTGCCAGCTGGCTACCGT 360	N E A Y R F R V A L P A Y P A S L T D V 120
	370	380	390	400	410	420	121	TCCCTGGCGCTGAGGCGCTGGCCGCCAACGACTCAGTATCTCGCTGAGTCAG 361	S L A L S E L R P N D S G I Y R C E V Q 140
	430	440	450	460	470	480	1421	CACGGCACTCGATGACAGCAGGCTGGAGCTCAAGTCAGTCAAGGTATCCACAGAC 480	H G I D D S S D A V E S S Q R Y P I Q T 160
	490	500	510	520	530	540	1481	CAACGAGGCCCTGTAACCGAACATGGAGCTTCGCCCGGGCTCCAGACATGGTG 540	P R E A C Y G D M D G F P G V R N Y G V 180

34/45

790 800 810 820 830 840
 781 CAGCGCTGTGGTGGGGCTTGCCTGGTGTCAAGACTCTCTCCCTTCCCCAACCCAGACT 840
 261 Q R C G G G L P G V K T L F L F P N Q T 280
 850 860 870 880 890 900
 841 GGCTTCCCCAATAAGCACAGCCGTTCAACGTCTACTGCTTCCGAGACTCGGCCAGCTT 900
 281 G F P N K H S R F N V Y C F R D S A Q L 300
 910 920 930 940 950 960
 901 CTGCCATCCCTGAGGCCTCCAACCCAGCCTCCAACCCAGCTTGATGGACTAGAGGCTATC 960
 301 L P S L R P P T Q P P T Q L D G L E A I 320
 970 980 990 1000 1010 1020
 961 GTCACAGTGACAGAGACCCCTGGAGGAAGTGCAGCTGCCTCAGGAAGCCACAGAGAGTGAA 1020
 321 V T V T E T L E E L Q L P Q E A T E S E 340
 1030 1040 1050 1060 1070 1080
 1021 TCCCGTGGGCCATCTACTCCATCCCCATGGAGGACGGAGGTGGAAGCTCCACT 1080
 341 S R G A I Y S I P I M E D G G G S S T 360

FIG.10D

35/45

1090 1100 1110 1120 1130 1140
 1081 CCAGAACCCAGCAGAGGCCCTAGGACGCTCCTAGAACATTGAAACACAATCCATGGTA 1140
 361 P E D P A E A P R T L L E F E T Q S M V 380
 1150 1160 1170 1180 1190 1200
 1141 CCGCCCACGGGTTTCAGAAGAGGAAGGTAAGGCATTGGAGGAAGAAGAGAAATATGAA 1200
 381 P P T G F S E E E G K A L E E E E K Y E 400
 1210 1220 1230 1240 1250 1260
 1201 GATGAAGAAGAGAAAGAGGAGGAAGAAGAAGAGGAGGAGGTGGAGGTGAGGCTCTGTGG 1260
 401 D E E E K E E E E E E V E D E A L W 420
 1270 1280 1290 1300 1310 1320
 1261 GCATGGCCCAGCGAGCTCAGCAGCCGGCCCTGAGGCCTCTCTCCCCACTGAGCCAGCA 1320
 421 A W P S E L S S P G P E A S L P T E P A 440
 1330 1340 1350 1360 1370 1380
 1321 GCCCAGGAGGAGTCACTCTCCAGGCAGCAAGGGCAGTCCTGCAGCCTGGTGCATCA 1380
 441 A Q E E S L S Q A P A R A V L Q P G A S 460

FIG.10E

FIG.11A

1390 1400 1410 1420 1430 1440
 1381 CCACTTCCCTGATGGAGAGTCATCGAACGCTTCCAGCTCCAGGTCATGGAACCACTTAC 1440
 461 P L P D G E S A S R P R V H G P R T 480
 1450 1460 1470 1480 1490 1500
 1441 GAGACTCTGCCACCTCCAGGAGAGCACTTACAGCACTCCACATCCACATCCACCTGGCT 1500
 36/45 1510 1520 1530 1540 1550 1560
 1501 GAGGCCAACGAGGTGGGGAGGCACTGGCTCTGTTCTGATGACTCTGGCTCCCTGGAGG 1560
 501 E A R E V G E A T G G P E L S G V P R G 520
 1561 GGGGCCGTTACCCAAATTCTGCCCTATAG 1587
 521 G A R T Q F A L * 528
 1 M S A D G A E A D G S T Q V T V E E P V 20
 1 ATGCTCTGCCAGGGCACTGGATGGCAAGCACCAAGTGGACAGTGGAAACCCGTA 60
 70 80 90 100 110 120
 61 CAGCACCCAGTGGTGGACCTGGCCAGCATGGCTCTGGCTCTGATCACGTCACCTGGAC 120
 21 Q Q P S V V D R V A S M P L I S T C D 40
 37/45 130 140 150 160 170 180
 121 ATGGTCCGCAAGCCTATGCCACCAAGGAGCTACCCGCACTGGACACTGTCGC 180
 41 M V S A Y A S T K E S Y P H V K T V C 60
 122 ATGGTCCGCAAGCCTATGCCACCAAGGAGCTACCCGCACTGGACACTGTCGC 180
 41 M V S A Y A S T K E S Y P H V K T V C 60
 181 GACGCCAGCAAGGAGTCAGGACCCCTCACGGCCCTGCTCACGGGCTCACGCCG 240
 61 D A A E K G V R T L T A A V S G A Q P 80

FIG.10F

1390 1400 1410 1420 1430 1440
 1381 CCACTTCCCTGATGGAGAGTCATCGAACGCTTCCAGCTCCAGGTCATGGAACCACTTAC 1440
 461 P L P D G E S A S R P R V H G P R T 480
 1450 1460 1470 1480 1490 1500
 1441 GAGACTCTGCCACCTCCAGGAGAGCACTTACAGCACTCCACATCCACATCCACCTGGCT 1500
 36/45 1510 1520 1530 1540 1550 1560
 1501 GAGGCCAACGAGGTGGGGAGGCACTGGCTCTGTTCTGATGACTCTGGCTCCCTGGAGG 1560
 501 E A R E V G E A T G G P E L S G V P R G 520
 1561 GGGGCCGTTACCCAAATTCTGCCCTATAG 1587
 521 G A R T Q F A L * 528
 1 M S A D G A E A D G S T Q V T V E E P V 20
 1 ATGCTCTGCCAGGGCACTGGATGGCAAGCACCAAGTGGACAGTGGAAACCCGTA 60
 70 80 90 100 110 120
 61 CAGCACCCAGTGGTGGACCTGGCCAGCATGGCTCTGGCTCTGATCACGTCACCTGGAC 120
 21 Q Q P S V V D R V A S M P L I S T C D 40
 36/45 130 140 150 160 170 180
 121 ATGGTCCGCAAGCCTATGCCACCAAGGAGCTACCCGCACTGGACACTGTCGC 180
 41 M V S A Y A S T K E S Y P H V K T V C 60
 122 ATGGTCCGCAAGCCTATGCCACCAAGGAGCTACCCGCACTGGACACTGTCGC 180
 41 M V S A Y A S T K E S Y P H V K T V C 60
 181 GAGACTCTGCCACCTCCAGGAGAGCACTTACAGCACTCCACATCCACCTGGCT 1500
 61 GAGACTCTGCCACCTCCAGGAGAGCACTTACAGCACTCCACATCCACCTGGCT 1500

250 260 270 280 290 300

241 ATCCTCTCCAAGCTGGAGCCCCAGATTGCATCAGCCAGCGAATACGCCACAGGGGGCTG 300
 81 I L S K L E P Q I A S A S E Y A H R G L 100

310 320 330 340 350 360

301 GACAAGTTGGAGGAGAACCTCCCCATCCTGCAGCAGCCCACGGAGAAGTCCTGGCGGACA 360
 101 D K L E E N L P I L Q Q P T E K S W R T 120

370 380 390 400 410 420

361 CAACGACTTGTGTCGTCTAAAGTGTGGGGCCCAAGAAATGGTGTCTAGGCCAACGACA 420
 121 Q R L V S S K V S G P K K W C L A P T T 140

430 440 450 460 470 480

421 CGGTGGCCACCAATTGTGGAGCGGTGGACGCCGACCCGCGGTGCTGTGCAGAGCGGCGTG 480
 141 R W P P I V G A V D A T R G A V Q S G V 160

490 500 510 520 530 540

481 GACAAGACAAAGTCCGTAGTGACCGGCAGCGTCCAATCGTCATGGGCTCCGCTTGGC 540
 161 D K T K S V V T G G V Q S V M G S R L G 180

FIG.11B

550 560 570 580 590 600

541 GGCACGAGGCTGAGTGGGTCGACACGGTGTGGGAAGTCGGAGGAGTGGCGGACAAAC 600
 181 G T R L S G V D T V L G K S E E W A D N 200

610 620 630 640 650 660

601 CACCTGCCCTTACGGATGCCGAACGGCCATGCCACATCCCTGGATGGCTTCGAC 660
 201 H L P L T D A E L A R I A T S L D G F D 220

670 680 690 700 710 720

661 GTCGCGTCCGTGCAGCAGCAGCGGAGGAACAGAGCTACTCGTACGTCTGGCTCCCTG 720
 221 V A S V Q Q R Q E Q S Y F V R L G S L 240

730 740 750 760 770 780

721 TCGGAGAGGCTGCGGAGCACGCCATGAGCACTCGCTGGCAAGCTTCGAGCCACCAAG 780
 241 S E R L R Q H A Y E H S L G K L R A T K 260

39/45

FIG.11C

FIG. 11E

441 E R T Q R A P V S I M Q * 452
 1321 GAGAGACCTCACAGCGGCTCCGCTCTATAATGCACTGA 1359

FIG. 11D

HAICH28 (SEQ ID NOS:23 and 24)

42/45

10	20	30	40	50	60
1 ATGGCGACCCAGCCTCGGCCAGACACACGGCTCTGGTGGCAGACTTGTAGGTTAT	60				
1 M A T P A S A P D T R A L V A D F V G Y	20				
70	80	90	100	110	120
61 AAGCTGAGGCAGAAGGGTTATGTCTGTGGAGCTGGCCCCGGGAGGGCCCAGCAGCTGAC	120				
21 K L R Q K G Y V C G A G P G E G P A A D	40				
130	140	150	160	170	180
121 CCGCTGCACCAAGCCATGCGGGCAGCTGGAGATGAGTCGAGACCCGTTCCGGCGCACC	180				
41 P L H Q A M R A A G D E F E T R F R R T	60				
190	200	210	220	230	240
181 TTCTCTGATCTGGCGGCTCAGCTGCATGTGACCCAGGCTCAGCCCAACAACGCTTCACC	240				
61 F S D L A A Q L H V T P G S A Q Q R F T	80				

FIG.12A

43/45

250	260	270	280	290	300
241 CAGGTCTCCGATGAACCTTTCAAGGGGGCCCAACTGGGGCCGCCTGTAGCCTCTTT	300				
81 Q V S D E L F Q G G P N W G R L V A F F	100				
310	320	330	340	350	360
301 GTCTTGGGGCTGCACTGTGTGCTGAGAGTGTCAACAAGGAGATGGAACCCTGGTGGGA	360				
101 V F G A A L C A E S V N K E M E P L V G	120				
370	380	390	400	410	420
361 CAAGTGCAGGAGTGGATGGCTACCTGGAGACGCGCTGGCTGACTGGATCCACAGC	420				
121 Q V Q E W M V A Y L E T R L A D W I H S	140				
430	440	450	460	470	480
421 AGTGGGGCTGGTTATCCCAGATCACTGAAGCTGAGATGGCTGATGAAGTAATTGCAGT	480				
141 S G G W L S Q I T E A E M A D E V I C S	160				
490	500	510	520	530	540
481 GAAATTAAAGCGACTGTGACTCTGCTGCAAGTTCCCCAGATCTGAGGAGCTGGAAGCT	540				
161 E I L S D C D S A A S S P D L E E L E A	180				

FIG.12B

FIG.12D

361 W Y S P Y * 365
1081 TGGTATTCCCTTACAA 1098

1090

341 S G F N S R P R G R V Y R A R A T S 360
1021 AGGGTATTAAACAGCAAGCCCGGCTGGCTCTAACAGGCGCCGGCTAGAGCACATCA 1080

1030 1040 1050 1060 1070 1080

45/45 321 P R A P Y R A R T T N Y N S S R F Y 340
961 CCAAGGCCGCTACCGGCCGACCAACCTACACAGCTCCGCTCTCGATTC 1020

970 980 990 1000 1010 1020

301 I K V I P K R T N R P G I S T D R G F 320
960 ATCAAGGTATCCCAAACAGAACACAGAACATCAAGAACACAGACCGGGTTT

910 920 930 940 950 960

281 D K E S V R T S L A L D E S L F R G R Q 300
841 GACAAAAGATGCTAGTGAAGACCTCCCTGGCCCTAGATGAGTCCTAGATTAGAGGAAAGGCCA 900

850 860 870 880 890 900

FIG.12C

261 T I L C D K F S G H P K G F A Y I E F S 280
781 ACCATCTGTTGACAAATTAGTGGCCATCCATTAGGTTGGCTATAGAGTC 840

790 800 810 820 830 840

241 G A T A E E L E A H F H G C G S V N R V 260
721 GTGCCACAGAGAGCTGGAAAGCTCACCTTCACTGGCTGCTGTCACGCTGT 780

730 740 750 760 770 780

44/45 221 S I E E K M E A D A R S I Y V G N V D Y 240
661 TCCATTGAGGAGAGATGAGCTGGATGCTCCATCTGCTGCAATGTTGGCAATGTTGAGAC 720

670 680 690 700 710 720

201 E V E K Q M N M S P P G N A G P V I M 220
601 GAGGTAGAGCACAGATGCTCACCTCCAGCAATGCTGGCCCGCTGATCA 660

610 620 630 640 650 660

181 I K A R V R E M E E A E K L K E L Q N 200
541 ATCAAAGCTCGAGTCAGGAGATGAGAGCTAGAGCTAAAGAGCTACAGAAC 600

550 560 570 580 590 600

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/12, 510, C07K 14/47, C12Q 1/68, A61K 3/817	A3	(11) International Publication Number: PCT/US98/00960	(10) International Publication Number: WO 98/31800
(22) International Filing Date: 21 January 1998 (21.01.98)		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(30) Priority Data: 60/034,204 60/034,205	21 January 1997 (21.01.97) 21 January 1997 (21.01.97)	US	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, UK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TI, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MA, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TI), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BI, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(21) International Application Number: PCT/US98/00960		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(22) International Filing Date: 21 January 1998 (21.01.98)	(43) International Publication Date: 23 July 1998 (23.07.98)
(71) Applicants (for all designated States except US): HUMAN GENOME SCIENCES, INC, 1 US/US; 940 Key West Avenue, Rockville, MD 20850 (US); AUCKLAND UNISERVICES LIMITED, INZ/NZ; Uniservices House, Level 7, 38 Symonds Street, Auckland 1001 (NZ).	
(72) Inventors and Inventors/Applicants (for US only): NI, Jian (CN/US); 5502 Manoffield Road, Rockville, MD 20853 (US); ROSEN, Craig, A. (US/US); 22400 Rolling Hill Road, Laytonsville, MD 20882 (US); GENTZ, Reiner, L. (DE/US); 13404 Fairland Park Drive, Silver Spring, MD 20904 (US); FENG, Ping (CN/US); 4 Reida Court, Gaithersburg, MD 20878 (US); KRISSANSEN, Geoffrey, W. (INZ/NZ); 157 B Grand Drive, St. Johns, Auckland 1001 (NZ); YARD, Jeffrey, Y. (CA/US); 443 West Side Drive #304, Gaithersburg, MD 20878 (US).	
(74) Title: HUMAN PROTEINS	

(57) Abstract

The present invention relates to novel human proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells and recombinant methods for producing the proteins of the invention. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT:	
AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BY	Belarus
BR	Burkina Faso
BU	Bulgaria
BN	Brunei
BR	Brazil
BY	Belarus
CA	Canada
CG	Central African Republic
CO	Colombia
CR	Costa Rica
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GL	Guinea
GN	Guinea
GR	Greece
GU	Guinea-Bissau
HU	Hungary
IE	Iceland
IL	Israel
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LC	Liberia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
ML	Mali
MM	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Szczecin
TD	Chad
UG	Uganda
TK	Togo
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Vietnam
YU	Yugoslavia
ZW	Zimbabwe

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT:

INTERNATIONAL SEARCH REPORT		INTERNATIONAL SEARCH REPORT	
<p>Internal Application No PCT/US 98/00960</p> <p>According to International Patent Classification (IPC) or to both national classification and IPC</p> <p>B. FIELDS SEARCHED</p> <p>Minimum documentation searched (classification system followed by classification symbols)</p> <p>IPC 6 C12N C07K C12Q A61K</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>Electronic data base consulted during the international search (name of data base and, where practical, search terms used)</p>		<p>Internal Application No PCT/US 98/00960</p> <p>C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT</p> <p>Category* Citation of document, with indication, where appropriate, of the relevant passage</p> <p>X T. FUJIWARA: "Otsuka cDNA project" EMBL SEQUENCE DATABASE, 4 October 1996, HEIDELBERG, FRG, XP002070248 Human placenta cDNA 5'-end GEN-548B01; Accession no. C17492; ---</p> <p>X H. FRIESS ET AL.: "A pancreatic cancer-specific expression profile" EMBL SEQUENCE DATABASE, 20 May 1996, HEIDELBERG, FRG, XP002070249 Human clone xs255 mRNA sequence; Accession no. U46281; ---</p> <p>A WO 92 04376 A (HEART RES INST) 19 March 1992 see the whole document ---</p>	
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <p>Category* Citation of document, with indication, where appropriate, of the relevant passage</p>		<p>Relevant to claim No.</p> <p>X 1,2,9-14</p> <p>X 1,2,9-14</p> <p>A 1-15</p>	
<p>D. DOCUMENTS CITED AS PRIOR ART</p> <p>Category* Citation of document, with indication, where appropriate, of the relevant passage</p>		<p>Relevant to claim No.</p> <p>A 1-15</p>	
<p>E. OTHER DOCUMENTS</p> <p>Further documents are listed in the continuation of box C.</p>		<p><input checked="" type="checkbox"/> Patient family members are listed in annex.</p>	
<p>* Special categories of cited documents :</p> <p>'A' document defining the general state of the art which is not considered to be of particular relevance</p> <p>'E' earlier document but published on or after the international filing date</p> <p>'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another document or other relevant fact (as specified)</p> <p>'O' document referring to an oral disclosure, use, exhibition or other means</p> <p>'P' document published prior to the international filing date but later than the priority date claimed</p>			
<p>Date of the actual completion of the international search</p> <p>2 July 1998</p>		<p>Date of mailing of the international search report</p> <p>09.0.98</p>	
<p>Name and mailing address of the ISA</p> <p>European Patent Office, P. B. 518 Patentlan 2 NL-2280-HV Rijswijk Tel. (+31-70) 340-2640, T. 31 651 0901 Fax: (+31-70) 340-2616</p>		<p>Authorized officer</p> <p>HORNIG H.</p>	

INTERNATIONAL SEARCH REPORT

In' tional application No.
PCT/US 98/00960

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see further information sheet

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

International Application No. PCT/US 98/00960

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: (1-15) partially

An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide selected from SEQ ID no.2; said nucleotide sequence which is at least 95% identical to a nucleotide sequence selected from SEQ ID no.1; an isolated nucleic acid molecule which hybridizes under stringent conditions to said nucleic acid molecule; a method of making a recombinant vector comprising said vector; a method of making an isolated polypeptide of SEQ ID no. 2; an isolated polypeptide of SEQ ID no.2; an isolated antibody capable of specifically binding to said polypeptide;

2. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 3 and 4;

3. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 5 and 6;

4. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 7 and 8;

5. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 9 and 10;

6. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 11 and 12;

7. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 13 and 14;

8. Claims: (1-15) partially

Item as subject 1 but limited to SEQ ID nos. 15 and 16;

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Information on patent family members				International Application No. PCT/US 98/00960
Patent document cited in search report	Publication date	Patient family member(s)	Publication date	
WO 9204376 A	19-03-92	AU 652575 B AU 8504091 A EP 0556506 A US 5731166 A	01-09-94 30-03-92 14-07-93 24-03-98	
EP 0263072 A	06-04-88	AU 7936987 A DE 3789413 D DE 3789413 T DK 517787 A ES 2052602 T F1 874287 A IE 61773 B JP 63157997 A JP 9188698 A PT 83849 B US 5702920 A US 5350687 A	28-04-94 28-07-94 04-04-88 16-07-94 04-04-88 30-11-94 30-06-88 22-07-97 31-07-90 30-12-97 27-09-94	
WO 9003394 A	05-04-90	FR 2636953 A AT 111920 T AU 6386882 B AU 4405689 A DE 68918399 D DE 68918399 T DK 130290 A EP 0367641 A ES 2053801 T JP 3502697 T	30-03-90 15-10-94 08-07-93 18-04-90 27-09-94 16-03-95 25-05-90 09-05-90 16-11-94 20-06-91	
WO 9517506 A	29-06-95	US 5527884 A AU 1441895 A CA 2178735 A EP 0749476 A JP 9506778 T	18-06-96 10-07-95 29-06-95 27-12-96 08-07-97	
WO 9503328 A	02-02-95	AU 7531294 A CA 2167693 A CN 1133049 A DE 4425481 A EP 0710248 A	20-02-95 02-02-95 09-03-96 02-03-95 08-05-96	

International Application No. PCT/ US 98/00960

FURTHER INFORMATION CONTINUED FROM PCT/ISAV 210

9. Claims: (1-15) partially

Idem as subject 1 but limited to SEQ ID nos. 17 and 18;
10. Claims: (1-15) partially

Idem as subject 1 but limited to SEQ ID nos. 19 and 20;
11. Claims: (1-15) partially

Idem as subject 1 but limited to SEQ ID nos. 21 and 22;
12. Claims: (1-15) partially

Idem as subject 1 but limited to SEQ ID nos. 23 and 24;

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/00960

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9503328 A		JP 9500531 T US 5770366 A ZA 9405278 A	21-01-97 23-06-98 19-01-96
WO 9014432 A	29-11-90	US 5580753 A AT 147436 T AU 637620 B	03-12-96 15-01-97 03-06-93
		AU 5928990 A CA 205597 A DE 69029657 D	18-12-90 24-11-90 20-02-97
		DK 473724 T EP 0473724 A ES 2099096 T JP 4506006 T	14-04-97 11-03-92 16-05-97 22-10-92
		US 5734037 A US 5414071 A	31-03-98 09-05-95
WO 9617925 A	13-06-96	AU 4639396 A CA 2205688 A FI 972390 A NO 972455 A	26-06-96 13-06-96 05-06-97 06-06-97
US 5536637 A	16-07-96	US 5712116 A	27-01-98

